The US Army's Center for Strategy and Force Evaluation
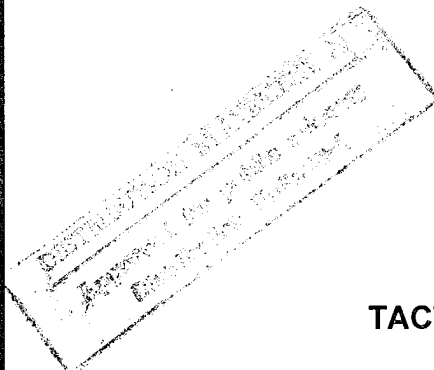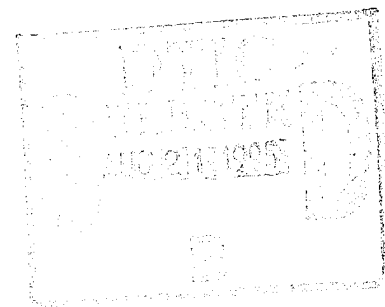
PERSONNEL ATTRITION RATES IN HISTORICAL LAND COMBAT OPERATIONS: SOME EMPIRICAL RELATIONS AMONG FORCE SIZES, BATTLE DURATIONS, BATTLE DATES, AND CASUALTIES

MARCH 1995

19950818 084

PREPARED BY
TACTICAL ANALYSIS DIVISION

US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

CONCEPTS ANALYSIS AGENCY

CAA-RP-95-1

REPORT DOCUMENTATION PAGE		*Form Approved* *OPM NO. 0704-0188*

Public reporting burden for this collection information is estimated to 1 hour per response, including the time for reviewing instructions, searching existing data sources gathering and maintaining the data needed, and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information. Including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of information and Regulatory Affairs, Office of Management and Budget, Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE **March 1995**	3. REPORT TYPE AND DATES COVERED **Final, Apr 94 - Feb 95**	
4. TITLE AND SUBTITLE **Personnel Attrition Rates in Historical Land Combat Operations: Some Empirical Relations Among Force Sizes, Battle Durations, ...**			5. FUNDING NUMBER
6. AUTHOR(S) **Dr. Robert L. Helmbold**			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **US Army Concepts Analysis Agency** **8120 Woodmont Avenue** **Bethesda, MD 20814-2797**			8. PERFORMING ORGANIZATION REPORT NUMBER **CAA-RP-95-1**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSORING/ MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILBILITY STATEMENT **Approved for public release; distribution unlimited**			12b. DISTRIBUTION CODE **A**

13. ABSTRACT (Maximum 200 words)

This paper uses historical data to develop some of the empirical relationships connecting the number of total battle casualties (TBC) to force sizes and battle durations. Here TBC is defined to be the sum of the killed in action (KIA), wounded in action (WIA), and captured or missing in action (CMIA). Because these relations may be affected by battle date, long-term trends with respect to battle date are also included.

14. SUBJECT TERMS **Personnel, attrition, combat, battle, casualties, losses, trends, history**			15. NUMBER OF PAGES *149*
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT **UNCLASSIFIED**	18. SECURITY CLASSIFICATION OF THIS PAGE **UNCLASSIFIED**	19. SECURITY CLASSIFICATION OF ABSTRACT **UNCLASSIFIED**	20. LIMITATION OF ABSTRACT **UL**

NSN 7540-01-280-5500
Standard Form 298

DEPARTMENT OF THE ARMY
U.S. ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

REPLY TO
ATTENTION OF:

0 9 AUG 1995

CSCA-TCT (5-5d)

MEMORANDUM FOR Deputy Under Secretary of the Army (OR), Headquarters, Department of the Army, Washington, DC 20310

SUBJECT: Personnel Attrition Rates in Historical Land Combat Operations: Some Empirical Relations Among Force Sizes, Battle Durations, Battle Dates, and Casualties

1. The U.S. Army Concepts Analysis Agency (CAA) is pleased to publish this research paper by Dr. Robert L. Helmbold. Its analysis of the principal data bases on personnel casualties over extended periods of time gives U.S. Army operations analysts a much improved foundation for judging future casualty numbers, casualty fractions, and casualty rates in historical land combat operations. Properly used, this information can be exploited to improve U.S. Army treatment of personnel attrition in models, wargames, studies, and analyses. Wide dissemination will make this work available to others for further use in their work.

2. Questions or inquiries should be directed to the Tactical Analysis Division, U.S. Army Concepts Analysis Agency (CSCA-TCT), 8120 Woodmont Avenue, Bethesda, MD 20814-2797, (301) 295-1611 or DSN 295-1611.

Encl

E. B. VANDIVER III
Director

PERSONNEL ATTRITION RATES IN HISTORICAL LAND COMBAT OPERATIONS: SOME EMPIRICAL RELATIONS AMONG FORCE SIZES, BATTLE DURATIONS, BATTLE DATES, AND CASUALTIES

March 1995

Prepared by

TACTICAL ANALYSIS DIVISION

US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, Maryland 20814-2797

PREFACE

The Personnel Attrition Rates (PAR) Study as a whole is limited to studying personnel strengths and battle casualties in historical land combat operations. Other types of attrition (nonbattle losses, losses to equipment, casualties to other services, and so forth) are outside PAR's scope, as are personnel losses in models, simulations, wargames, field experiments, or training exercises (like those of the National Training Center).

Phase 1, or PAR-P1, was devoted to assembling the available data and past studies on personnel strengths and attrition rates in land combat operations, preparing a comprehensive bibliography of it, and planning the approach to subsequent phases. Its specific objectives were to:

- Collect as many as possible of the available tabulated data and data-based studies of attrition rates in historical land combat operations,
- Prepare a comprehensive bibliography of such data and studies, and
- Outline an approach to accomplishing the subsequent phases of the PAR Study as a whole.

The bibliography of works collected during Phase 1 was published as *Personnel Attrition Rates in Land Combat Operations: An Annotated Bibliography,* US Army Concepts Analysis Agency Research Paper, CAA-RP-93-2, June 1993 (AD-A268 787). The collection of data and data-based studies consists of the files of pertinent documents maintained at the US Army Concepts Analysis Agency.

Phases 2 and 3 of the PAR Study will convert some of the most important data to electronic form in order to facilitate its analysis and will perform selected analyses of the attrition data to derive information useful in US Army wargames, studies, and analyses. As of this writing, the following documents have been published during Phase 2:

- *Personnel Attrition Rates in Historical Land Combat Operations: Susceptibility and Vulnerability of Major Anatomical Regions,* CAA Research Paper CAA-RP-93-3, August 1993, AD-A270 766.
- *Personnel Attrition Rates in Historical Land Combat Operations: A Catalog of Attrition and Casualty Data Bases on Diskettes Usable With Personal Computers,* CAA Research Paper CAA-RP-93-4, September 1993, AD-A279 069 (report), AD-M000 344 (diskettes).
- *Personnel Attrition Rates in Historical Land Combat Operations: A Note on the Probability of Readmissions and Multiple Wounds,* CAA Research Paper, CAA-RP-94-2, April 1994.

This paper, written as part of Phase 2, furnishes an additional analysis. It uses historical data to develop some of the empirical relationships connecting the number of total battle casualties (TBC) to force sizes, battle durations, and battle dates. Here TBC is defined to be the sum of the killed in action (KIA), the wounded in action (WIA), and the captured or missing in action (CMIA). Because these relations may be affected by battle date, long-term trends with respect to battle date are also included.

The basic approach used was to review the prior work in this area and then to analyze the available data bases for information related to long term trends in personnel attrition. The focus is on the analysis of the general trends in and relations among force sizes, battle durations, and casualties. Our efforts seek to advance the state of the art over prior efforts by (i) giving the Constant Fallacy appropriate recognition, (ii) using a regression model that includes the battle duration and battle date as potentially important factors, (iii) employing robust regression to minimize the distorting effects of a few gross errors in the data, (iv) systematically using more than one data base at a time in order to determine the sensitivity of the results to different sets of data, and (v) using several dependent variables, to include the casualty numbers as well as the casualty exchange ratio. The primary data analysis technique used is descriptive statistics.

ACKNOWLEDGEMENTS

	PERSONNEL ATTRITION RATES IN HISTORICAL LAND COMBAT OPERATIONS: SOME EMPIRICAL RELATIONS AMONG FORCE SIZES, BATTLE DURATIONS, BATTLE DATES, AND CASUALTIES	SUMMARY CAA-RP-95-1

THE REASON FOR PERFORMING THIS RESEARCH was that the estimation of attrition in future combat engagements might be improved if any general relationships connecting casualties with force sizes and battle durations can be discovered and exploited.

THE SPONSOR was the Director, US Army Concepts Analysis Agency.

THE OBJECTIVE was to search for some empirical relationships among force sizes, battle durations, and battle casualties in historical land combat operations, using the comprehensive bibliography and data base collection previously assembled in the Personnel Attrition Rates (PAR) studies.

THE SCOPE OF THE RESEARCH is restricted to consider mainly total battle casualties (TBC), defined to be the sum of its principal components, namely, the killed in action (KIA), the wounded in action (WIA), and the captured or missing in action (CMIA).

THE MAIN ASSUMPTION of this paper is that the bulk of the pertinent works has been collected and is on file at CAA and that statistical procedures are appropriate for summarizing the empirical relationships inherent in these data. A secondary assumption, needed for application of the findings, is that the statistics of future battles will be like the statistics of past battles; in other words, that trends of sufficiently long duration can be extrapolated to the near future with a reasonable degree of confidence.

THE BASIC APPROACH used in this study is to analyze the available data bases for information related to long-term trends in personnel attrition. The primary technique used is descriptive statistics.

THE PRINCIPAL FINDINGS of the work reported herein are that personnel casualty numbers, casualty fractions, and casualty rates have declined with the passage of time over the last 400 years or so. On the average, this decline appears to have proceeded at a fairly predictable rate. However, the scatter of data about the average is large, and so the average trend is only a rough indication of the level of personnel casualties that might be experienced in future combat. Also, the loser has tended to suffer substantially larger casualty fractions and casualty rates than the winner.

THE STUDY EFFORT was directed by Dr. Robert L. Helmbold, Tactical Analysis Division.

CONTENTS

PREFACE

TABLES

CHAPTER 1

EXECUTIVE SUMMARY

1-1. BACKGROUND. In April 1992, the US Army Concepts Analysis Agency (CAA) started a three-phased study of personnel attrition data—the Personnel Attrition Rates (PAR) Study. PAR as a whole is limited to studying personnel strengths and battle casualties in historical land combat operations. Other types of attrition (nonbattle losses, losses to equipment, casualties to other services, and so forth) are outside PAR's scope, as are personnel losses in models, simulations, wargames, field experiments, or training exercises (like those of the National Training Center).

Phase 1, or PAR-P1, was devoted to assembling the available data and past studies on personnel strengths and attrition rates in land combat operations, preparing a comprehensive bibliography of it, and planning the approach to subsequent phases. Its specific objectives were to:

- Collect as many as possible of the available tabulated data and data-based studies of attrition rates in historical land combat operations,
- Prepare a comprehensive bibliography of such data and studies, and
- Outline an approach to accomplishing the subsequent phases of the PAR Study.

The survey of works collected during Phase 1 was published as *Personnel Attrition Rates in Land Combat Operations: An Annotated Bibliography*, US Army Concepts Analysis Agency Research Paper, CAA-RP-93-2, June 1993 (AD-A268 787). The collection of data and data-based studies consists of the files of pertinent documents maintained at the US Army Concepts Analysis Agency.

In Phases 2 and 3 of the PAR Study, some of the most important data are being converted to electronic form in order to facilitate its analysis, and some selected analyses of it will be performed to derive information useful in US Army wargames, studies, and analyses. As of this writing, the following documents have been published during Phase 2:

- *Personnel Attrition Rates in Historical Land Combat Operations: Susceptibility and Vulnerability of Major Anatomical Regions,* CAA Research Paper CAA-RP-93-3, August 1993, AD-A270 766.
- *Personnel Attrition Rates in Historical Land Combat Operations: A Catalog of Attrition and Casualty Data Bases on Diskettes Usable With Personal Computers,* CAA Research Paper CAA-RP-93-4, September 1993, AD-A279 069 (report), AD-M000 344 (diskettes).
- *Personnel Attrition Rates in Historical Land Combat Operations: A Note on the Probability of Readmissions and Multiple Wounds,* CAA Research Paper, CAA-RP-94-2, April 1994.

The present paper, written as part of Phase 2, furnishes an additional analysis. It uses historical data to develop some of the empirical relationships connecting the number of total battle casualties (TBC) to force sizes and battle durations. Here TBC is defined to be the sum of the killed in action (KIA), the wounded in action (WIA) and the captured or missing in action (CMIA). Because these relations may be affected by battle date, long-term trends with respect to battle date are also included. These long-term trends are expressed primarily in terms of the TBC number, the TBC fraction (defined as the ratio of the TBC number to the size of the force), and the TBC rate (defined as the ratio of the TBC fraction to the duration of time over which the TBC were inflicted—usually expressed as the number of TBC per 1,000 personnel-days).

1-2. OBJECTIVE. The objective of this research paper is to examine the historical evidence for long-term trends in the general relationships among force sizes, battle durations, force ratios, casualty numbers, casualty fractions, and casualty rates, and thereby to establish a baseline for projections into the future.

1-3. SCOPE

a. PAR as a whole is limited to studying personnel strengths and battle casualties of land combat forces. Other types of attrition (nonbattle losses, losses to equipment, casualties to other services, and so forth) are outside PAR's scope. PAR is concerned only with historical data on actual combat operations; it will not deal with personnel losses in models, simulations, wargames, field experiments, or training exercises (like those of the National Training Center). PAR focuses mainly on either original or translated works in English, although some important work in other languages may be included. Studies of personnel attrition are also included, provided they contain cogent analyses of a publicly available, nonproprietary body of tabulated data on attrition in actual combat operations. Since trends in attrition over long periods of time are of interest, data on ancient as well as recent battles are solicited. However, as no contract support is anticipated and in-house resources are limited, no systematic effort is made to extract data from the archives or primary source materials, and no original historical research is envisioned. Thus, PAR relies almost exclusively on secondary works that contain data in readily usable tabulated form. All works received prior to the cutoff date of 31 May 1994 are included.

b. The issues to be examined in this paper are grouped into two general groups, as listed below. They were gleaned from a variety of sources. Each of these general groups is analyzed in its own chapter. These chapters list more specific issues whose resolution would illuminate that group's general issue.

- Group 1–What empirical trends in force sizes, battle durations, force ratios, casualty numbers, casualty exchange ratios, casualty fractions, and fractional exchange ratios of the opposing sides persisted over extended periods of time?

- Group 2–How are force sizes, battle durations, force ratios, casualty numbers, casualty fractions, and casualty rates interrelated?

c. Additional issues, which we hope to examine in future works, include such items as the following.

- Group 3–How are force sizes, battle durations, force ratios, casualty numbers, casualty fractions, and casualty rates related to winning and losing?

- Group 4–How are force sizes, battle durations, force ratios, casualty numbers, casualty fractions, and casualty rates related to various situational and environmental factors, such as the rate of advance, nationality, tactics, terrain, and supporting fires, among others?

- Group 5–How do casualty numbers, casualty fractions, and casualty rates vary over relatively brief periods of time?

- Group 6–What proportion of the total battle casualty number are due to killed in action, wounded in action, died of wounds, captured, and missing in action?

- Group 7–How are force sizes, battle durations, force ratios, casualty numbers, casualty fractions, and casualty rates distributed statistically?

- Group 8–What other questions would we like to address?

1-4. ASSUMPTIONS. The main assumptions of this paper are (i) that the bulk of the pertinent works has been collected and is on file at CAA and (ii) that statistical procedures are appropriate for summarizing the empirical relationships implicit in these data. A secondary assumption, needed for application of the findings, is that the statistics of near-future battles will be like the statistics of the battles of the past 400 years or so—in particular, that trends of sufficiently long duration can be extrapolated to the near future with a reasonable degree of confidence.

1-5. APPROACH. The basic approach used in this study was to review the prior work in this area and then to analyze the available data bases for information related to long-term trends in personnel attrition. We focused on the analysis of the general trends in and relations among force sizes, battle durations, and casualties. Our efforts seek to advance the state of the art over prior efforts by (i) giving the Constant Fallacy (Helmbold-1994) appropriate recognition, (ii) using a regression model that includes the battle duration and battle date as potentially important factors, (iii) employing robust regression to minimize the distorting effects of a few gross errors in the data, (iv) systematically using more than one data base at a time in order to determine the sensitivity of the results to different sets of data, and (v) using several dependent variables, to include the casualty numbers as well as the casualty exchange ratio. The primary data analysis technique used is descriptive statistics.

1-6. FINDINGS AND OTHER OBSERVATIONS. The following are applicable to the period from 1600 AD to the present. Since they have persisted for a long period of time despite major changes in tactics and weaponry, they presumably can be extrapolated to the near future with a fair degree of confidence:

a. Battle durations have tended to increase.

b. Attacker and defender strengths have been fairly stable over time and tended to be nearly equal. The force ratio favoring the defender has been fairly stable over time, and defenders typically fought at a slight numerical disadvantage.

c. Attacker and defender TBC casualty numbers have declined over time and tended to be nearly equal. The casualty exchange ratio favoring the defender appears to have been fairly stable and close to unity.

d. Attacker and defender TBC casualty fractions have declined over time, and the defender's TBC fraction has tended to be greater than the attacker's. The fractional exchange ratio favoring the defender has been relatively stable over time.

e. Winner and loser strengths exhibit different trends with different data bases, some data bases showing an increase and others either a decrease or no appreciable change. However, all the data bases agree that the force ratio favoring the winner has been stable and close to unity.

f. Winner and loser TBC casualty numbers have declined over time, with the loser's casualties typically at least twice those of the winner. The casualty exchange ratio favoring the winner has been more or less stable over time, depending on the data base used.

g. Winner and loser TBC casualty fractions have declined over time, with the loser's casualty fraction typically at least twice that of the winner. The fractional exchange ratio favoring the winner has been fairly stable over time.

h. Some of the trends differ from one data base to another. However, all of the data bases agree that strength is not particularly associated with victory in battle and that the casualty or fractional exchange ratio are incomparably more strongly associated with victory in battle. From past research (Helmbold-1986), it would seem that the fractional exchange ratio is a somewhat better index of victory in battle than the casualty exchange ratio.

i. It is not true that the defender has some inherent advantages over the attacker. In fact, the attacker has generally taken fewer TBC than the defender. On the average, the casualty exchange ratio favoring the defender is less than 1, and the fractional exchange ratio favoring the defender is also less than 1.

j. Smaller forces take and inflict proportionately more casualties than larger forces. It is conjectured that this is a result of diminishing returns to scale.

k. Neither Lanchester's square law, Osipov's law, nor Peterson's logarithmic law are good approximations to the true relation of casualty fractions to force ratios.

l. The data do not reveal the expected (that is, approximately linear) dependency of casualty numbers on the temporal duration of a battle. One might expect that the casualty numbers would be in direct proportion to the duration of a battle, but this is clearly not what the data show.

m. The casualty exchange ratio favoring the defender *(CERY) decreases* as the force ratio favoring the defender *(FRY) increases*, despite what one might have expected. In fact, the approximate equation relating *CERY* to *FRY* is

$$CERY \approx \frac{1}{2}(FRY)^{-0.4}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

n. However, the fractional exchange ratio favoring the defender *(FERY)* does increase as the force ratio favoring the defender increases. The approximate equation relating *FERY* to *FRY* is

$$FERY \approx \frac{1}{2}(FRY)^{+0.6}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

o. The casualty exchange ratio favoring the winner *(CERW) decreases* as the force ratio favoring the winner *(FRW) increases*, despite what one might have expected. In fact, the approximate equation relating CERW to FRW is

$$CERW \approx 7.4(FRW)^{-0.5}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

p. However, the fractional exchange ratio favoring the winner *(FERW)* does increase as the force ratio favoring the winner *(FRW)* increases. The approximate equation relating *FERW* to *FRY* is

$$FERW \approx 7.4(FRW)^{+0.5}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

q. Other casualty relations appear to differ, depending on the data base use, the battle duration, and the battle date. The reasons for these differences are left to future investigations.

CHAPTER 2

APPROACH

2-1. INTRODUCTION. This chapter outlines the approach taken to analyze various issues regarding personnel casualties and attrition in land combat operations.

2-2. DEFINITIONS. This paragraph introduces terminology that is used consistently throughout the remainder of this paper. Since the terminology in general use is often imprecise or ambiguous, it is necessary to clarify our use of the following terms.

    a. (Personnel) Casualties (or Casualty Number). The term "casualties" will often be used in the general (but ambiguous) sense to refer to any of a number of personnel casualty notions, such as the number, fraction, or rate of personnel casualties. In addition, the unqualified term "casualties" will sometimes be used when it is clear from the context that it refers to the number of personnel casualties taken by a force during the course of a particular action or time period. However, we always use the precise technical expression "number of casualties" or "casualty number" to stand for the number of personnel casualties taken by a force during the course of a particular action or time period. We use the notation C for the number of casualties and add postfixes to identify the force that suffered those casualties. Thus, CX will be the attacker's casualty number, CY the defender's, CW the winner's, and CL the loser's. The symbol Z will be used as a generic symbol for the side. Thus, CZ is the number of casualties to side Z, where Z may take on any of the values $X, Y, W,$ or L. Throughout this paper, casualties are expressed in units of persons. For example, if in battle Bravo the defending forces suffered 24,000 casualties, then $CY = 24,000$ persons.

    b. (Personnel) Strength and Force Ratio. By the (initial personnel) strength of a side we mean its strength at the start of a particular action or engagement. Force size is a synonym for strength. We use the notation $X0$ for the attacker's (initial personnel) strength, and $Y0$ for the defender's. The postfix zero is a reminder that this represents the initial strength. In this paper, strength is expressed in units of persons, so that, for example, if in battle Bravo the defending force's strength was 100,000, then $Y0 = 100,000$ persons.

    The force ratio is defined as the ratio of one side's strength to that of the other. More precisely stated, the force ratio favoring side Z is defined to be the ratio $FRZ = Z0 / \overline{Z}0$, where $\overline{Z}0$ is the size of the force opposing side Z.

    c. (Personnel) Casualty Fraction. By the casualty fraction we mean the casualty number of a force, expressed as a proportion of its size. We use the notation F, with appropriate suffixes or postfixes, for casualty fractions. By definition, $FZ = CZ / Z0$, where Z may be any of the symbols identifying the force under consideration (e.g., Z may be X, Y, etc.). Casualty fractions are physically "dimensionless," that is, they are pure numbers. In this paper they are expressed as fractions or percentages. Thus, if in battle Bravo the defender's casualties are

$CY = 24,000$ and the defender's strength is $Y0 = 100,000$, then the defender's casualty fraction is $FY = CY / Y0 = 0.24$, or 24 percent.

d. (Battle) Duration. The duration of a battle is the time that elapses between its beginning and end, or the total time during which actual combat took place (as opposed to pauses for maneuvering, regrouping, etc.). In this paper, we adopt the day as the standard unit of duration, which is usually symbolized by T. Thus, a battle that lasts 10 hours has a duration of $T = 10 / 24 = 0.417$ days.

e. (Personnel) Casualty (or Attrition) Rate. By the casualty (or attrition) rate we mean the average rate at which the (personnel) casualty fraction has increased during a particular action or period of time. We use the notation R, with appropriate suffixes or postfixes, for the casualty rate or attrition rate. By definition, $RZ = FZ / T \equiv CZ / (Z0 \times T)$, where T is the duration of the action or period under consideration and Z identifies which force is under consideration. In this paper, we express the resultant attrition rate as the number of casualties per thousand personnel-days. Thus, if the defender's casualty fraction is $FY = 0.24$ for an action that lasted 3 days, then $RY = FY / T = 0.24 / 3 = 0.080$ per person-day, or 80 per thousand person-days. The convenient abbreviation "/kpd" is used as an abbreviation for the phrase "per thousand person-days," so the last result is written as "80/kpd."

f. Casualty Exchange Ratio. The casualty exchange ratio favoring side Z is defined to be

$$CERZ = \frac{C\overline{Z}}{CZ}.$$

g. Fractional Exchange Ratio. The fractional exchange ratio favoring side Z is defined to be

$$FERZ = \frac{F\overline{Z}}{FZ} \equiv \frac{C\overline{Z}}{CZ} \frac{Z0}{\overline{Z}0} \equiv CERZ \times FRZ.$$

h. Battle Date. The battle date is the date on which the battle commenced, expressed in years Anno Domini and fractions thereof. Thus, a battle that started on 1 July 1890 would have a battle date of 1890.500. In much of our later work, we found it convenient to convert these dates to a centered and scaled version defined by

$$D \equiv DateCent = \frac{(BattleDate - 1800)}{100}.$$

In this form, the battle dates are centered at 1800 AD and normalized to an elapsed date unit of 1 century. In this centered and scaled version, a battle that began on 1 July 1890 would have a normalized and scaled battle date of $D = 0.905$.

2-3. DISCUSSION OF THE DEFINITIONS. Frequently, the available data are such that we must be content with approximations to the quantities defined above, or may need to modify them to handle complicated situations. The following explains how some particular cases are handled.

a. The first issue is what type of casualties is included in C. In this paper, we usually are concerned with total battle casualties, defined as the sum of the killed in action, wounded in action, and captured or missing in action. When other types of casualties are used, this will be stated explicitly. To emphasize this we sometimes write *TBCX* for the attacker's TBC number, *TBCFW* for the winner's TBC fraction, *TBCERY* for the TBC casualty exchange ratio favoring the defender, and so forth.

b. The next issue is whether--and how--to adjust the initial force strength when a force is either augmented by personnel reinforcements and/or replacements, diminished by detachments, or is altered by any other exogenous personnel changes. Here, "exogenous" is intended to indicate the operation of factors other than those that directly affect TBC. When there are exogenous changes, the initial strength can be a poor index of the effective force size, and it would be desirable to adjust it in some way to account for these personnel changes. For most of the data bases used in this paper, no such adjustment is feasible because the requisite data are not provided. As explained later, we elected to use only data on initial strengths, rather than to adjust the data for exogenous changes.

However, we provide the following to illustrate one apparently reasonable approach to adjusting force strengths when the requisite data are available. This approach adopts the view that the adjustment should be chosen to retain the notion that multiplying the (effective initial personnel) strength by the temporal duration of the action should give the total personnel effort (in person-days) committed to the action. Accordingly, when exogenous changes in strength occur, we take the (effective initial personnel) strength to be given by the average number of person-days of effort expended during the course of the action or time period under consideration, ignoring the reductions caused by battle casualties.

For example, if a force begins battle Bravo with 70,000, but is reinforced by 60,000 at the end of 4 hours and has 30,000 detached by higher command at the end of 10 hours, and if the battle lasts a total of 12 hours, then the (effective initial personnel) strength of the attacker is taken to be

$$\{70{,}000 \times 4 + (70{,}000 + 60{,}000) \times 6 + (70{,}000 + 60{,}000 - 30{,}000) \times 2\} \div 12 = 105{,}000,$$

the time average of the total number of personnel-hours that were committed to this action, ignoring battle casualties.

c. The last case to consider is how to define the duration of a battle that consists of several phases, each of varying intensity. For example, suppose battle Bravo starts at 0800 and continues with relatively high intensity to 1200. Action is then suspended, only to be resumed at 1600 and finally concluded at 1800. One could argue that the battle lasted either 0.25 days (4 hours from 0800 to 1200, plus 2 hours from 1600 to 1800), or 0.417 days (10 hours elapsed from

the start at 0800 to the conclusion at 1800), or 1 day (since all the action took place within a single day). In order to distinguish among them, we call the first duration the active duration, the second the elapsed duration, and the last the overall or whole day duration. Statistics based on these various durations will be identified similarly—that is, as active, elapsed, and overall (whole day) statistics. In this paper we attempt to provide statistics based on each of the time periods for which we have suitable data. However, we often know only the dates on which the battle took place, and in such cases can provide only overall (whole day) statistics.

In this paper the standard unit of time is the day. Thus, we express attrition rates in units of casualties per 1,000 (effective) initial strength per day, whether the duration used is the active, overall, or whole day duration. Note that expressing attrition *rates* in units of days does not imply that the time *duration* under consideration is actually a whole day. For example, a force with an (effective) initial strength of 1,000 personnel that suffers an active or overall casualty rate of 100/kpd has not necessarily lost 100 persons, nor was it necessarily in action for a full day. Indeed, if that force was in action for 1/10 of a day, then it actually lost only 10 persons during the course of that action, because {10 casualties}÷{(1,000 initial strength)×(1/10 of a day duration)}×{1,000 personnel per kpd} = 100/kpd.

2-4. ISSUES TO BE EXAMINED. The issues to be examined in this paper can be considered as falling into two general categories, as listed below. They were gleaned from a variety of sources, including Berndt-1897, Bodart-1908, Smith-1955, Helmbold-1961, Helmbold-1969, Helmbold-1971, Helmbold-1986, Dupuy-1990, Helmbold-1993, and others. Each of these general groups is analyzed in its own chapter. These chapters list more specific issues whose resolution would illuminate that group's general issue.

- Group 1—What empirical trends in force sizes, battle durations, force ratios, casualty numbers, casualty exchange ratios, casualty fractions, and fractional exchange ratios of the opposing sides persisted over extended periods of time?

- Group 2—How are force sizes, battle durations, force ratios, casualty numbers, casualty exchange ratios, casualty fractions, and fractional exchange ratios interrelated?

2-5. SELECTION OF DATA BASES

a. The work in this paper requires data bases with information on the strengths and losses of both sides in several battles. Four main data bases with such information were selected for use in the analyses presented in this paper. They appear to be practically the only ones available that are suitable for our analyses. Where appropriate, derived data bases that use only a portion of one or another of the main data bases were used. This section is devoted to describing the four main data bases used in this paper.

b. The following main data bases were selected for use in this paper.

(1) CDB91DAT. This data base is essentially the same as the CDB90-1991 data base described in detail in Helmbold-1993, with the following corrections for typographical errors:

- ISEQNO 205 (Cold Harbor), change FINSTA from 96707 to 96907.

- ISEQNO 242 (The Yalu), change FINSTA from 54900 to 54890.

- ISEQNO 338 (West Wood I), change INTSTA from 1748 to 1940.

- ISEQNO 371 (Sommepy Wood), change FINSTA from 9081 to 8610.

- ISEQNO 371 (Sommepy Wood), change FINSTD from 217 to 220.

It contains 660 battles ranging in date from 1600 to 1982. However, for the work in this paper, the battles were filtered to omit all that had reinforcements or replacements. Thus, only a fraction of the total number of battles is actually used in this paper. None of the other data bases used in this paper distinguish between initial forces and any exogenous changes to it.

(2) BWSHALL. This data base is essentially the same as the BWSH-1993 data base described in detail in Helmbold-1993. It contains 1097 battles ranging in date from 1619 to 1905. However, some of these battles are identified as sieges. In this paper, the sieges were filtered out and only the nonsiege battles were used.

(3) BODASHIP. This data base is essentially the same as the BODASHIP-1993 data base described in detail in Helmbold-1993. It contains 120 naval battles ranging in date from 1638 to 1905. It has been included as a sort of check on how far the findings based on the land combat data bases may generalize to naval and other battles.

(4) PARCOMBO. This data base consolidates into a single spreadsheet the information in the PARMISC-1993, SP128-1961, SP190-1964, and DODGE-1993 data bases, each of which is described in detail in Helmbold-1993. However, when this was done, it was noticed that there were several duplicate entries. Moreover, in some cases, these duplicate entries gave significantly different numbers for strengths and/or losses, and/or differed with regard to which side was attacking or which side won. The following procedure was used to resolve such conflicting items of information.

(a) First, only one version of the data was carried forward to the PARCOMBO data base. In case of duplicates, the version carried forward was checked against the reference materials cited below for the best supported figures on strengths and losses, for which side was attacking, and for a determination of which side won. In addition, these sources were used to provide whole day durations for the battles in the SP128-1961 and SP190-1964 data bases. The references used were Harbottle-1905, Eggenberger-1967, Laffin-1986, and Dupuy-1970.

(b) Second, for those DODGE-1993 battles which were not duplicates, missing information on which side attacked and on which side won was also obtained by reference to the references cited in the immediately preceding paragraph.

(c) The upshot of this process was the creation of the PARCOMBO data base, with 368 battles ranging in date from 280 BC to 1965 AD. However, only a handful of these battles is

from dates earlier than 1500 AD. For the work described in this paper, we filtered out all of the PARCOMBO battles with dates earlier than 1500 AD.

c. Several other data bases are listed in Helmbold-1993 but are not used in this paper for various reasons. In some cases, the data bases provide data on only one side, while we require data on both sides. Others contain data on wars, while in this paper we consider only battles. Others deal with air battles, but the force sizes involved in them are so much smaller than those for the land or naval battles that any comparison might miss the point. Hence, for this paper, we decided to use only the four major data bases, CDB91DAT with 660 battles ranging in date from 1600 AD to 1982 AD, BWSHALL with 1087 battles ranging in date from 1619 AD to 1905 AD, BODASHIP with 120 naval battles ranging in date from 1638 AD to 1905 AD, and PARCOMBO with 368 battles ranging in date from 280 BC to 1965 AD.

d. There is some overlap between the three land battle data bases in regard to the specific battles included (especially for the battles of the Thirty Years' War, the wars of Frederick the Great, the Napoleonic war battles, and the battles of the American Civil War), and it must be presumed that there is a moderately high correlation among them for the numbers reported on those battles (since, presumably, they were all drawing on the same basic source material). However, there is also a considerable degree of difference among them in regard to the specific battles included, and each source came to a semi-independent judgment on the numbers. In short, it seems reasonable to treat these data bases as being quasi-independent samples of the battles from military history.

e. As far as the reliability of these data bases is concerned, our personal judgment is that the CDB91DAT data base is the best and most reliable of the three and that the others are rather less reliable. Nevertheless, they all contain errors and peculiarities that tend to distort and confuse the general trend or average result. Accordingly, methods of analysis should be selected to minimize the consequences of gross errors. The following paragraphs illustrate the practical difficulties in establishing accurate numbers for the personnel engaged or lost in battles.

(1) Order of General Robert E. Lee, as quoted by William F. Fox in Fox-1889:

HEADQUARTERS ARMY OF NORTHERN VIRGINIA

General Orders, No. 63.   May 14, 1863

The practice which prevails in the Army of including in the list of casualties those cases of slight injuries which do not incapacitate for duty, is calculated to mislead our friends, and encourage our enemies, by giving false impressions as to the extent of our losses.

The loss sustained by a brigade or regiment is by no means an indication of the service performed or perils encountered, as experience shows that those who attack most rapidly, vigorously, and effectually generally suffer the least. It is, therefore, ordered that in future the reports of the wounded shall only include those whose injuries, in the opinion of the medical officers, render them unfit for duty. It has also been observed that the published reports of casualties are in some instances accompanied by a statement of the number of men taken into action. The commanding general deems it unnecessary to do more than direct the attention of officers to the impropriety of thus furnishing the enemy with the means of computing our strength, in order to insure the immediate suppression of this pernicious and useless custom.

By command of General Lee   W. H. Taylor, *Assistant Adjutant-General*

(2) The following are selected observations on data quality from HERO-1967:

(a) (p vi) "Information on strengths and casualties was derived from US sources entirely for selected operations in Okinawa and Korea, and from German as well as US records for selected operations in the European Theater in World War II. Where records were inadequate or ambiguous, available figures have been expanded or modified on the basis of professional military and historical judgment. (This was particularly necessary for German data, since all of the most relevant German records have been returned to West Germany without having been microfilmed.)"

(b) (p 2) "Enemy records now available in the United States rarely include data for units below division level, and those for US combat elements available for this study were limited in quality and level of resolution by the purpose for which they were originally compiled.

"Because of the nature of warfare as well as the nature of the records, it has proved impossible to provide meaningful figures on an hourly basis. Some information is available on the intensity and duration of combat on certain days, but it is no more possible to ascertain the distribution of casualties by category or time than it is valid to assume that they were spread evenly over the duration of the combat."

(c) (p 4-5) "No Japanese, North Korean, or Chinese Communist records are available. Consequently, figures on casualties and strengths of those forces were procured entirely from the reports of the opposing US units, and must be viewed with considerable caution.

"In only a few instances were daily reports of German strengths and losses found, and these covered isolated periods of a few days. Much of the material was in the form of monthly reports at the corps or army level. In some cases information pertaining to the same period was found in different forms, although frequently conflicting, necessitating evaluation, and application of professional judgment."

(d) (p 6) "In the calculations and analyses made of the data the average unit strength during the engagement was normally used. When daily strengths were not available, figures representing the strength at each end of the period, normally a month, were usually at hand, and from these an average daily figure was derived. In cases where the only German casualty figures available were those accumulated for a stated period, usually ten days or one month a daily breakdown of casualties was estimated, based upon knowledge of the situation existing, the nature of the combat in which the units were engaged throughout the period, the intensity of the combat indicated by casualty figures for US forces, knowledge of the course of operations, and experience with similar forces in similar situations."

(e) (p 7) "Japanese casualty figures on Okinawa, derived solely from US sources, included only killed, broken down into several categories, including estimated dead as well as counted dead and estimated numbers sealed in caves ... the accuracy of these daily figures is impossible to validate. There are no figures at all on Japanese wounded ... it was found that doubling the number of counted dead, while ignoring other estimated categories, gave the most plausible total for dead and wounded."

(f) (p 7) "For Korea, ... we have accepted the Far East Command (FEC) figures ....
These evidently do not include estimates of wounded. We have assumed they do."

(g) (p A-4) "While no detailed casualty reports of 16th Panzer Division for the entire
period were found, total casualties of 1,300 for the division for the Salerno battle were reported
by XIV Panzer Corps to Tenth Army. ... These casualties have been somewhat arbitrarily broken
down according to the known postures of the division at various times during the period, using
professional judgment on the basis of the situation, the known circumstances, and the casualties
of the 45th Division during the same periods. Thus, it has been assumed that not more than 380
of the German casualties were suffered during the last 16 days, when the action was less intense,
and of these not more than 20 were suffered in the last ten days (through October 5). This leaves
a total of 920 casualties to be allocated to the first 6 days of the operation. Fifty casualties have
been allocated to each of the first two days, and 100 casualties to the next day. This leaves 720
casualties for the period of intensive counterattack, September 12-14. Despite the arbitrariness of
these allocations, the orders of magnitude must be reasonable correct, and the results of using
these estimates for the purposes of the study will not be significantly distorted from what must
have actually occurred." [Its not entirely clear why the orders of magnitude "must" be correct, or
how distorted the results might be from what actually occurred.]

(h) (p A-6) "Microfilmed records of XIV Panzer Corps contained several sets of
strength and casualty figures for 26th Panzer Division, all of them unfortunately at some variance
with each other."

(i) (p A-8) "The scattered German records ... included scattered daily reports of
casualties, three 10-day summations of casualties, and one report of total casualties by type and
unit for the period November 4-15. Taken together, they indicated a total of 892 battle casualties
in November, of which 433 could be confidently attributed to 17 of the 30 days involved. The
balance of 396 was then prorated over the remaining 13 days on the basis of the nature and
intensity of combat as revealed by tactical accounts."

(j) (p A-12) "Three separate statements of 45th Infantry Division strength and
casualties were found, each of which was in considerable disagreement with the others. Under
the necessity of making an arbitrary choice, ... that with the most recent date was accepted. The
nonavailability of detailed German strength and casualty figures was particularly frustrating in
attempting to analyze results of 45th Division operations at Anzio. No strength figures for the
German divisions engaged in whole or in part against the division were discovered for February.
... Thus an approximation of each division's strength was arrived at by multiplying the
ascertainable infantry strength in each instance by a factor of 5. ... A comparison of the results of
this process with secondary historical works ... suggests that the overall German strength was
probably somewhat higher than that produced by this method. ... For this reason, in each of the
Anzio engagements considered in detail, an arbitrary 25% has been added to the strengths of the
German units positively identified, and 25% has also been added to their casualty totals."

(k) (p A-13) "Casualties of the 45th Infantry Division were reported ... the most
recently dated document was used, although there were indications that, in some cases, casualty

figures reflected accumulations of casualties of previous days which were reported late. In these cases, clearly erroneous totals have been redistributed among prior days on the basis of known intensity of combat."

(l) (p A-14) "Total [German] casualties were broken down to correspond with fractions of divisions engaged against the 45th Infantry Division ...."

(m) (p C-8) "The data available for the 7th and 96th Divisions was quite complete, insofar as strength and casualties were concerned. In the previous research, which had made these data available to HERO, however, there had been no need to record either strengths or casualties of units attached to, or directly supporting, these divisions. Accordingly, the casualty figures derived from analysis of the various key engagements of these divisions is applicable only to the division strength. On the other hand, the availability of the supporting forces is essential in order to be able to evaluate the overall opposing strength ratios, and the amount of force and firepower applied in order to inflict the casualties on opposing Japanese forces in the division sectors."

(n) (p D-6) "The considerations and procedures for development and analysis of North Korean and Chinese Communist strengths and casualties are similar to those for Okinawa, but with some differences. In the first place, opposing overall force strengths and structures are not so well known as was the case on Okinawa. Similarly, the casualties are based mainly on estimates."

2-6. SPECIFIC METHODOLOGICAL PRACTICES. We attempt to adhere to the following specific methodological practices as best we can.

a. Issues must be expressed in precise (preferably mathematical) form. Otherwise, they are much too vague to be meaningful. Expressions to the effect that, "Generally, the attacker has more casualties than the defender" are too ambiguous to be meaningful. For example, is this to be interpreted as stating that the casualty numbers of the attacker are higher than those of the defender in 99 percent of all battles? Higher than those of the defender in 51 percent of all battles? Or is the statement intended to refer to casualty fractions rather than to casualty numbers? And how much "higher" is meant? Does it mean that the attacker's casualty numbers are 10 percent higher than the defender's in 51 percent of all battles, and lower in the remaining 49 percent? Does it mean that the attacker's casualty numbers are treble those of the defender in 99 percent of all battles? When it comes to testing an issue, we must know precisely which version of it is being tested.

b. When a number of specific issues are under consideration, we try to formulate a single (but more general) issue that embraces all of them, yet is sufficiently precise to be tested quantitatively. When such a general but precise issue can be found, its quantitative analysis serves to resolve all of the issues raised by the collection of more specific issues.

c. Untangling the effects of a number of simultaneously acting factors upon a final result is a general scientific problem. Unfortunately, no entirely satisfactory resolution of this problem has as yet been found. However, for initial efforts—such as this, the general practice is to assume that the factors operate independently, or (at worst) additively, until proven otherwise.

d. We adopt the following, which we dub "Richardson's Principle," from page 132 of Richardson-1960, emphasis in the original. "This discussion of change in history has an important influence on all the rest of the book. *Let us assume, as a working hypothesis, that every finite set of historical events is only a sample of what might have happened. Any quantitative theory of history is therefore not required to agree precisely with actual historical events but to agree only within the range of uncertainty ascribable to sampling.* In particular cases there may be much difficulty in ascertaining the appropriate range of sampling; nevertheless, the above principle directs the inquiry."

e. The following can be cited in support of Richardson's Principle, as given in the preceding paragraph:

(1) "For want of a nail, the shoe was lost. For want of a shoe, the horse was lost. For want of a horse, the rider was lost. For want of rider, the battle was lost. For want of a battle, the kingdom was lost." Traditional saying.

(2) "I returned and saw under the sun, that the race is not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all." Ecclesiastes, 10:11.

(3) "Nothing is more subject to chance than war Clausewitz-1976.

CHAPTER 3

**ISSUE GROUP 1:
EMPIRICAL TRENDS OF FORCE SIZES,
BATTLE DURATIONS, AND CASUALTIES**

3-1. INTRODUCTION. This chapter is devoted to the general question, "What empirical trends in force sizes, battle durations, force ratios, casualty numbers, casualty exchange ratios, casualty fractions, and fractional exchange ratios of the opposing sides persisted over extended periods of time?"

3-2. BACKGROUND. In this paragraph we mention briefly the chief works known to us on the empirical trends of force sizes, force ratios, battle durations, and casualties in order to indicate our present understanding of them. However, we do not attempt a comprehensive review or critique of the current state of affairs. Instead, our review of prior works is highly focused. We present only their principal results, and even within that narrow focus we seek only to capture the one or two points of greatest relevance to the present chapter. Accordingly, the reader should not be misled into thinking that the papers reviewed contain *only* the information we cite. On the contrary, each contains many important observations and insights worthy of attentive reading and thoughtful consideration, but which happen not to be related directly to our immediate interests.

    a. That the fractions of casualties in battles were not increasing with the passage of time was well known at least as early as Berndt-1897 and Bodart-1908. At the start of the 1900s, they both predicted that the casualty fraction would not increase with the passage of time. As we shall see, those predictions have been borne out by the experience of the last century.

    b. For example, Berndt-1897 (p 149) notes that "It is evident that not only the total losses but also the bloody losses have diminished [over the course of time], ... Therefore, battles have become less productive of losses, less lethal in the course of time." He also observes (p 158) that "The remarkable correspondence of this result to the earlier results obtained by historical considerations is the basis for my claim that in the major battles of a future war the average bloody loss to both sides will certainly not exceed 15% and even the bitterest ones fought will only with difficulty exceed 20% on both sides.

    "Although one can object that what was true earlier nowadays no longer applies, since at that time one had only muzzle loaders (later breech loaders); however, today armies are equipped with rapid fire repeater small arms in addition to smokeless powder, rapid fire cannon, and so forth.

    "Now, these objections immediately collapse when one realizes that major advances have previously been made in weapons technology without these altering in any way the—one can almost say—lawful ebbing of the losses. And yet the superiority of the breech loader over the muzzle loader was doubtless one of many valuable steps forward, as was the introduction of the

repeater rifle over the breech loader. Therefore, if it were a reasonable assumption that the repeater rifle would in the great majority of battles produce more devastation than its preceding small arm, which in turn was far better than its predecessor, the muzzle loader, then the breech loader should have been able to surpass its predecessor in producing battle results.

"The superiority of a new and better weapon strikes terror only when the opponent possesses only the older, poorer one, as was the case in 1866 [Prussian-Austrian War]. Nowadays this cannot occur, for all great armies have small arms of nearly equal quality, and hence the advantage is the same on both sides, *i.e.*, it raises both sides equally.

"Therefore, one need have no hope or fear that improvements in weapons, by the time they have become common goods, will produce greater results or higher losses; the implacable facts of history speak against it."

Berndt-1897 also observes (p 145) that "Regarding the duration of battles ... the striking observation is that, on the average, the duration of a battle is increasing with the passage of time, although one might have expected the opposite, considering the extraordinary improvements in weaponry and their corresponding vastly increased destructive action."

c. Bodart-1908 (p 41) says: "The percentage of bloody losses to armies has varied widely over the last four hundred years and up to the last great east-Asian [Russo-Japanese] war exhibited a noticeable tendency to decrease." He also says (p 43) "Wars of more recent times have become in the main less murderous than those of Napoleon. The Russo-Turkish War 1828-1829, the Russo-Polish War 1830-1831, and the US Civil War 1861-1865 exhibit 14% bloody losses; the Austrian-Italian 1848-1849 as well as the contemporary Hungarian Insurrection only 4%; the battles of the Crimean War (1853-1856) 12%; the Italian War of 1859 9.5%, of 1866 8%; the Franco-German War 1870-1871 (average of 20 battles) 7.5%; the Russo-Turkish War 1877-1878 about 14%; while in the civil war of 1899-1901 the proportion sank to 5%."

d. Smith-1955 apparently was the next to address this issue. He agrees with Berndt-1897 and Bodart-1908 that casualties in battles have declined with the passage of time, so that the average battle has become less intense. However, Smith-1955 suggests that battles have tended to occur more frequently in the course of campaigns and wars, so that the overall probability of a soldier becoming a casualty in a war has not changed much with the passage of time. Thus, they say (p 10): "The statement, often made that earlier wars were more deadly than modern ones, can be misleading. The average percentage of the total force who were killed or died of wounds each month during the major wars of the period considered [1750 to 1950] shows no pronounced trend. The average risk of becoming a casualty which a soldier faced, at any time during these wars, has remained approximately constant. Perhaps the explanation for this is that the rate at which Army Commanders were willing, or could afford, to accept casualties has not varied greatly. Thus, if the deadliness of wars is measured by the average risk, faced by the participants, of becoming a casualty at any time, it is not possible to say that wars have become markedly more or less deadly.

"It is true that before the present century heavy casualties were often taken in battles which were usually of short duration, but major battles were relatively infrequent, averaging less than one a month for the wars before World War I. Moreover, only about 10% of the effective army strength, on an average, was used in these battles. By comparison, daily casualty rates in modern wars do not normally reach such a high level, but the fighting tends to be continuous, merely fluctuating in intensity throughout the campaign."

e. Working independently of those mentioned before, Helmbold-1961 and Helmbold-1964, concluded that casualty fractions in battles have remained relatively stable, while battle durations have increased. The net result of these two trends is a decline in attrition rates for battles. It might be conjectured from this that less intense battles (that is, those with lower attrition rates) can continue for a longer period of time than more intense ones.

f. Much later, Dupuy-1990 popularized these findings. He notes (p 26) that "Despite the fact that weapons have become more lethal, the battlefield has rather steadily become less deadly over these same four centuries [1600 to the present]." He echoes Smith-1955's caution about extrapolating from battles to campaigns or to wars, observing that (p 39) "However, simply because casualty rates have been declining fairly steadily over the past 400 years does not mean that war has become either less dangerous or less horrible. ... Prior to the 20th Century, battles usually lasted only for one day or less, and there were periods of days, weeks, and months between battles. In the 20th Century, particularly during World War I, troops have been exposed to hostile fire in battles that continued day after day. The fact that daily battle casualty rates have been lower during the past century has been offset by the fact that these lower daily losses have been sustained day after day on a continuous basis."

3-3. SPECIFIC ISSUES TO BE ADDRESSED. In this chapter we address the following specific issues, all of which relate to long-term trends in battle casualties and related quantities. Any extrapolation of the findings from battles to campaigns or wars should bear in mind the cautions voiced by Smith-1955, and by others.

- Battle durations have increased over the last 400 years or so.
- The strengths of the opposing sides have increased over the last 400 years or so.
- Force ratios have stayed about the same since 1600 or so.
- Casualty numbers, casualty fractions, and casualty rates have increased with the passage of time, due to the increased lethality and effectiveness of the weapons employed.
- Casualty exchange ratios have stayed about the same since 1600.
- Attrition rates have tended to remain stable or decline for the last 300 or 400 years, despite the advances in weapons technology.
- Attrition rates have steadily and steeply declined with the passage of time since about 1600 AD.
- Fractional exchange ratios have stayed about the same since 1600.

3-4. LONG-TERM TRENDS IN BATTLE DURATIONS

a. Figures 3-1 through 3-4 show plots of durations (in days) versus battle date for the CDB91DAT, PARCOMBO, BWSHALL, and BODASHIP data bases, respectively. Here and throughout this paper, CDB91DAT includes only those battles with no reinforcements or replacements, PARCOMBO includes only battles that occurred later than 1500 AD, and BWSHALL includes only battles not identified as sieges.

b. For CDB91DAT, three different kinds of durations can be identified, as follows:

(1) Active durations are those obtained by summing the durations of each individual active period identified as having occurred during the course of the battle.

(2) Elapsed durations are the times from start of the first active period to the end of the last active period that occurred during the course of the battle.

(3) Overall durations are the durations in whole days.

For example, a battle that began at 0800 on 1 January, and that had an active period from 0800 to 1200 on that day, followed by an inactive period between 1200 and 1400, and concluded with an active period from 1400 to 1600 would have an active duration of 0.25 days (*i.e.*, 6 hours, comprised of 4 hours in the morning plus 2 in the afternoon), an elapsed duration of 0.33 days (*i.e.*, 8 hours, from 0800 to 1600), and an overall duration of 1 day (because the entire battle took place within the span of a single day).

c. Only overall durations are given in the data for most of the PARCOMBO battles. However, some of them provide information on shorter durations (such as "... the battle lasted for 9 hours"). We have not been able to determine whether these durations correspond most nearly to the active or to the elapsed durations, which are quite clearly defined for the CDB91DAT battles. They will be referred to as active durations, merely in order to have a name for them. The BWSHALL and BODASHIP data bases contain only overall durations.

d. Figure 3-1 shows that CDB91DAT durations increased with battle date, with the trend line being essentially the same for both the active and elapsed durations. Here, as in most other cases, the trend lines are exponential fits (that is, they plot as straight lines on log-linear coordinates). These trend lines are intended only to guide the eye to the general trend of the data. Figure 3-2 shows that PARCOMBO active durations also increased with battle date, and in very much the same way as for the CDB91DAT active and elapsed durations. Figure 3-3 shows that the BWSHALL overall durations also tend to increase with battle date. However, this data base has so many battles assigned a nominal overall duration of one day that the trend is masked. That is, if the less than one day active durations had been available, the trend would have appeared to be much steeper. From Figure 3-4, it appears that BODASHIP battle durations declined slightly, though rather insignificantly, with battle date. Here again, the use of nominal 1-day overall durations tends to mask a possible underlying upward trend.

3-5. LONG-TERM TRENDS IN ATTACKER AND DEFENDER STRENGTHS AND FORCE RATIOS. Figures 3-5 and 3-6 show the trends in attacker and defender strengths with

battle date. There seem to be only slight declines in attacker and defender strengths for the CDB91DAT and PARCOMBO battles. Observe that this, taken together with the increase in durations noted in the preceding paragraph, suggests that the amount of effort in personnel-days has tended to increase with battle date. Figure 3-7 shows that for the CDB91DAT battles the force ratio favoring the defender *(FRY)* has tended to be less than 1 (*i.e.*, the defender typically is at a numerical disadvantage), but has changed only marginally over time. Since both sides take part in a battle for the same length of time, the ratio of their efforts in personnel-days is the same as that of their force ratios. Figure 3-8 appears to show a noticeable decline in *FRY* for PARCOMBO battles.

3-6. LONG-TERM TRENDS IN ATTACKER AND DEFENDER CASUALTIES AND CASUALTY EXCHANGE RATIOS. Figures 3-9 and 3-10 show the trend of attacker and defender casualties. The trends for the CDB91DAT and PARCOMBO battles is decidedly toward substantially lower casualty numbers for both sides. Note that, although the CDB91DAT values drop faster than the PARCOMBO ones, the extrapolation to the year 2000 would give about the same result. Figures 3-11 and 3-12 show the trend in casualty exchange ratio favoring the defender (*CERY*). It appears that *CERY* has increased slightly for the CDB91DAT battles but has been essentially unchanged—or even decreased slightly—for the PARCOMBO ones.

3-7. LONG-TERM TRENDS IN ATTACKER AND DEFENDER CASUALTY FRACTIONS AND FRACTIONAL EXCHANGE RATIOS. Figures 3-13 and 3-14 show the trend of attacker and defender casualty fractions. The trend for the CDB91DAT battles is definitely toward lower casualty fractions for both sides. The PARCOMBO trends show a decline for the attacker, but an increase for the defender. Figures 3-15 and 3-16 show the trend in fractional exchange ratio favoring the defender (*FERY*). It appears that *FERY* has declined slowly for the CDB91DAT and at about the same rate for the PARCOMBO battles.

3-8. LONG-TERM TRENDS IN WINNER AND LOSER STRENGTHS AND FORCE RATIOS

   a. Figures 3-17 through 3-20 show the trend of winner and loser strengths. The trend is down for the CDB91DAT, PARCOMBO, and BODASHIP battles, but up for the BWSHALL battles. However, in each case, the trend line for the loser's strength is very nearly the same as for the winner's strength. This seems to imply that strength has little to do with victory in battle. It is interesting that three of the four data bases seem to show a cross-over somewhere around 1700 to 1750. If this cross-over is real (that is, not an accidental feature or artifact) its significance has yet to be analyzed and understood.

   b. Figures 3-21 through 3-24 show the trend of force ratio favoring the winner (*FRW*). The trend is a very gradual increase for all four data bases.

3-9. LONG-TERM TRENDS IN WINNER AND LOSER CASUALTIES AND CASUALTY EXCHANGE RATIOS

a. Figures 3-25 through 3-28 show the trend of winner and loser casualties. The trend is steeply downward for the CDB91DAT and PARCOMBO battles but is practically unchanged for the BWSHALL battles. The BODASHIP data show a different pattern, with the loser's casualties gradually declining and the winner's casualties steeply declining with the passage of time.

b. Figures 3-29 through 3-32 show the trend of casualty exchange ratio favoring the winner *(CERW)*. The trend is gradually downward for the CDB91DAT battles, gradually upward for the PARCOMBO battles, gradually downward for the BWSHALL battles, and steeply upward for the BODASHIP battles.

3-10. LONG-TERM TRENDS IN WINNER AND LOSER CASUALTY FRACTIONS AND FRACTIONAL EXCHANGE RATIOS

a. Figures 3-33 through 3-36 show the trend of winner and loser casualty fractions. The trend is steeply downward for the CDB91DAT battles. The PARCOMBO battles show a similar downward trend for the winner, but an gradual increase for the loser. The BWSHALL data have a gradually downward trend for the winner, and a steeper one for the loser, a pattern which is shared by the BODASHIP battles.

b. Figures 3-37 through 3-40 show the trend of fractional exchange ratio favoring the winner *(FERW)*. The trend is slightly downward for the CDB91DAT and BWSHALL battles. However, the PARCOMBO and BODASHIP battles show a rising trend. Note that (with the exception of some of the earlier PARCOMBO battles) the trend lines in each case are substantially greater than unity. This suggests that the FERW value is strongly associated with victory in battle. Moreover, this association appears to hold for naval battles as well as for land battles.

3-11. SUMMARY OF FINDINGS. The findings regarding long-term trends with respect to battle date are that:

- Battle durations have tended to increase.

- Attacker and defender strengths have been fairly stable over time, and tended to be nearly equal. The force ratio favoring the defender has been fairly stable over time, and defenders typically fought at a slight numerical disadvantage.

- Attacker and defender TBC casualty numbers have declined over time and tended to be nearly equal. The casualty exchange ratio favoring the defender appears to have been fairly stable and close to unity.

- Attacker and defender TBC casualty fractions have declined over time, and the defender's TBC fraction has tended to be greater than the attacker's. The fractional exchange ratio favoring the defender has been relatively stable over time.

- Winner and loser strengths exhibit different trends with different data bases, some data bases showing an increase and others either a decrease or no appreciable change. However, all the data bases agree that the force ratio favoring the winner has been stable and close to unity.

- Winner and loser TBC casualty numbers have declined over time, with the loser's casualties typically at least twice those of the winner. The casualty exchange ratio favoring the winner has been more or less stable over time, depending on the data base used.

- Winner and loser TBC casualty fractions have declined over time, with the loser's casualty fraction typically at least twice that of the winner. The fractional exchange ratio favoring the winner has been fairly stable over time.

3-12. CONCLUSIONS AND OBSERVATIONS. Some of the trends differ from one data base to another. However, all of the data bases agree that strength is not particularly associated with victory in battle, and that both the casualty and the fractional exchange ratio are incomparably more strongly associated with victory in battle. A more refined analysis of the matter (Helmbold-1986), has shown that the fractional exchange ratio is a decidedly better index of victory in battle than the casualty exchange ratio.

CDB91DAT: Durations vs Date
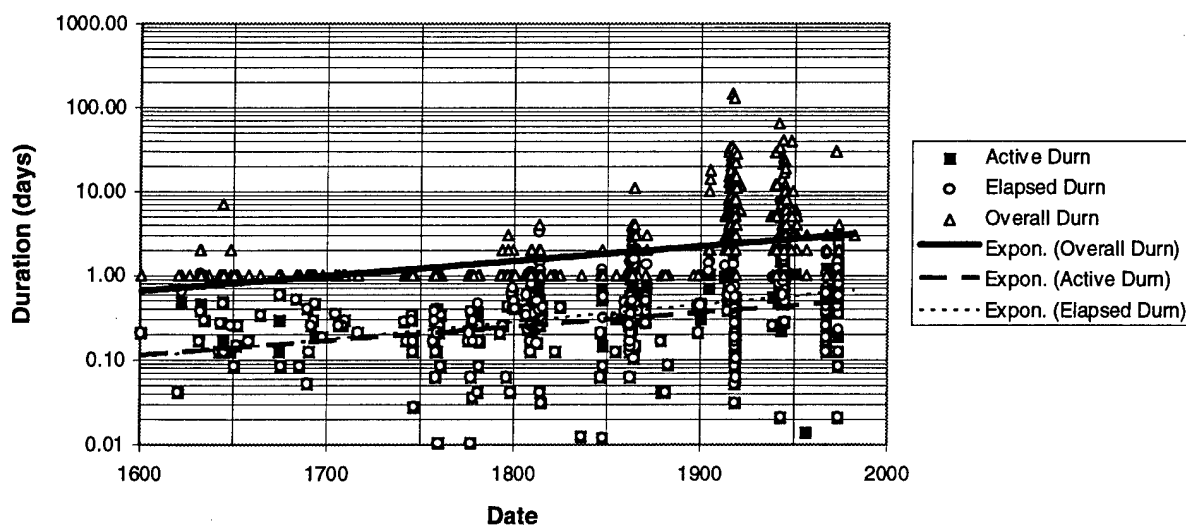
Figure 3-1. Duration versus Date for the CDB91DAT Data Base

PARCOMBO: Durations vs Date
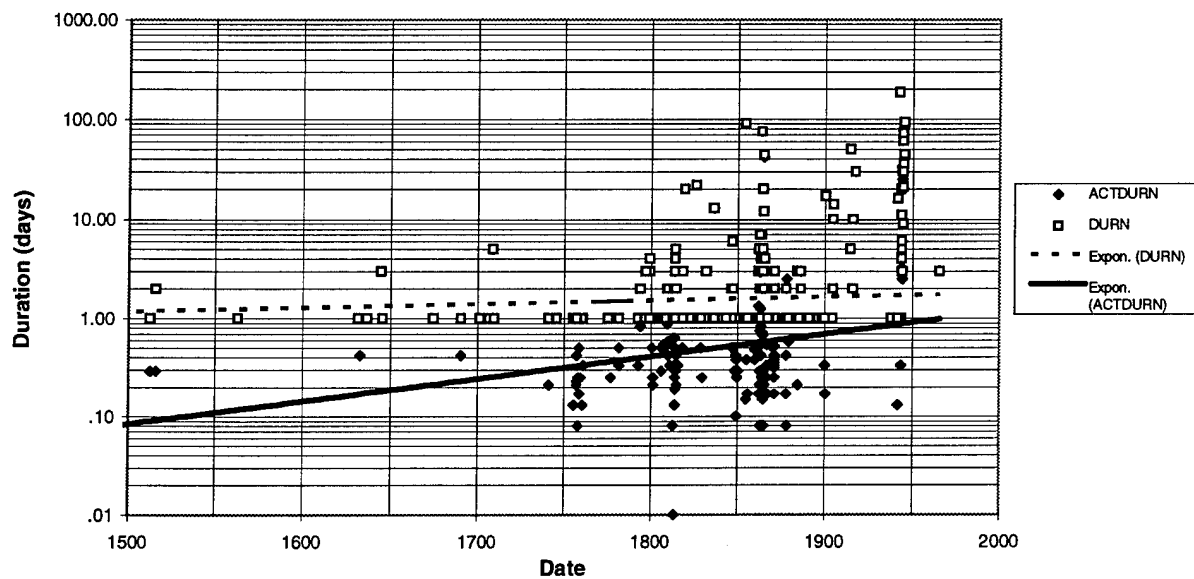
Figure 3-2. Duration versus Date for the PARCOMBO Data Base

BWSHALL: Duration vs Date
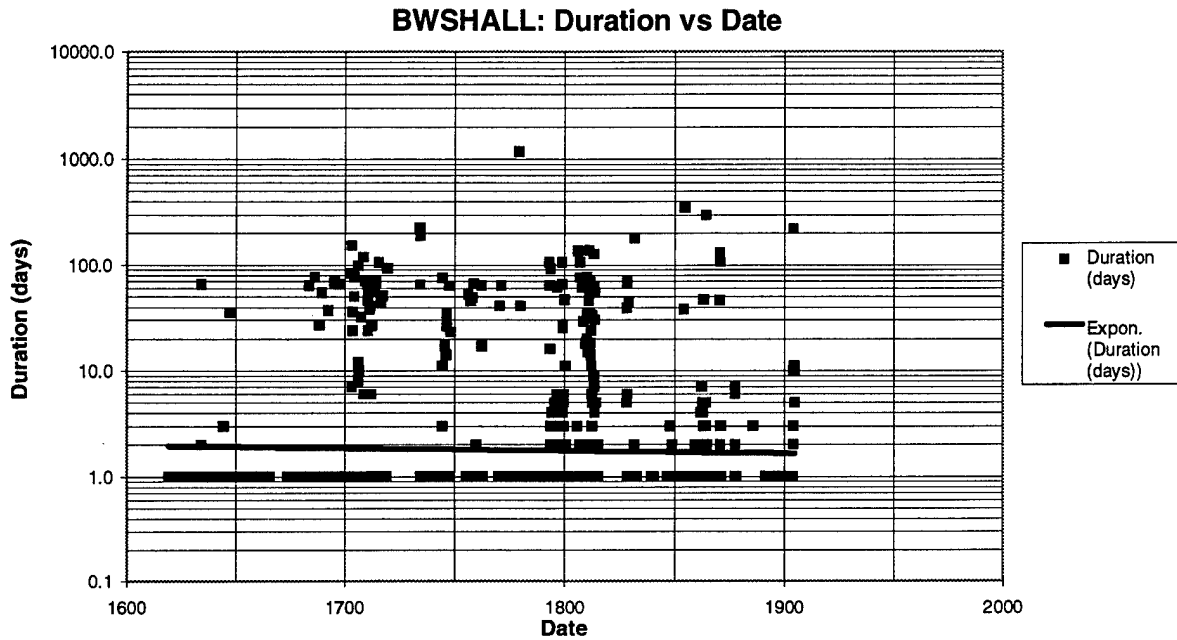
Figure 3-3. Duration versus Date for the BWSHALL Data Base

BODASHIP: Duration vs date
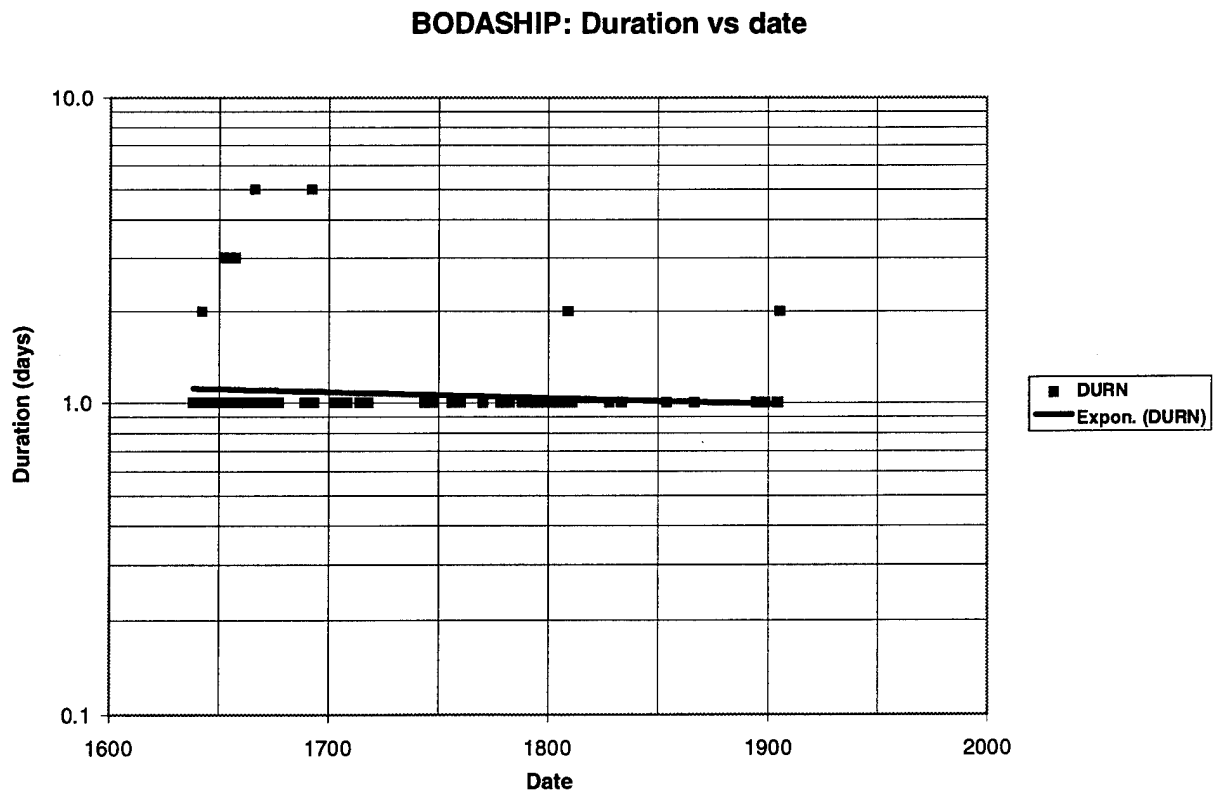
Figure 3-4. Duration versus Date for the BODASHIP Data Base

CDB91DAT: Strengths vs Date
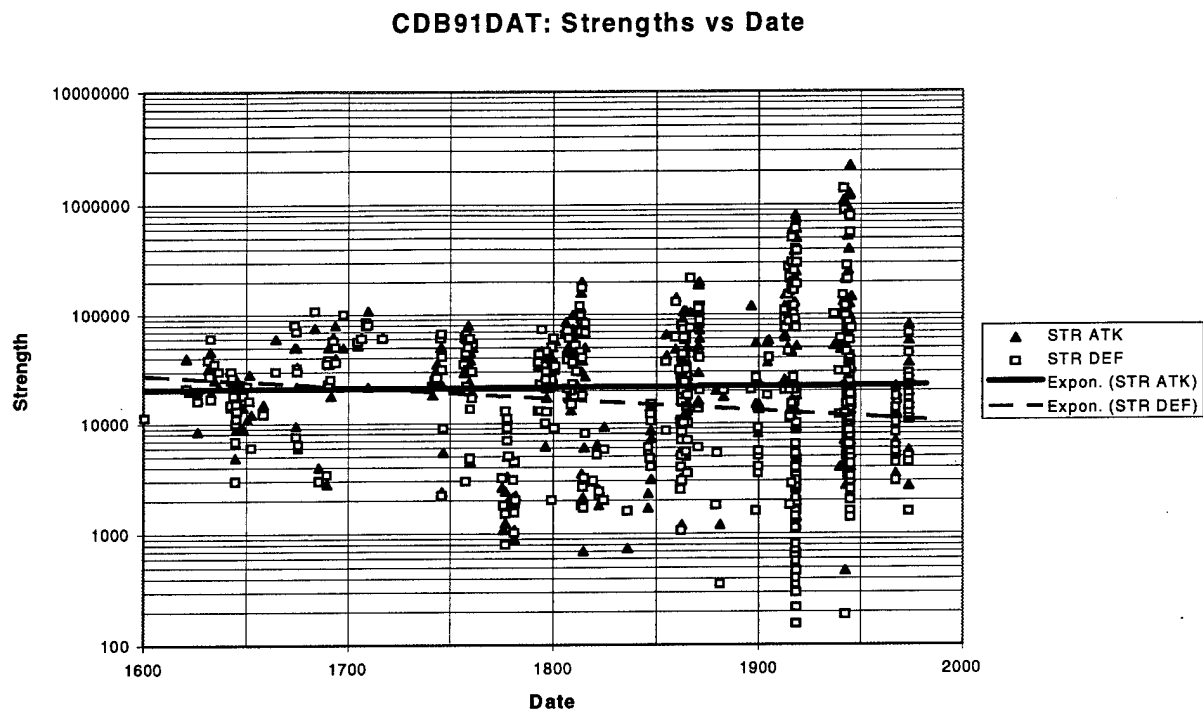
Figure 3-5. Attacker and Defender Strengths versus Date for the CDB91DAT Data Base

PARCOMBO: Strengths vs Date
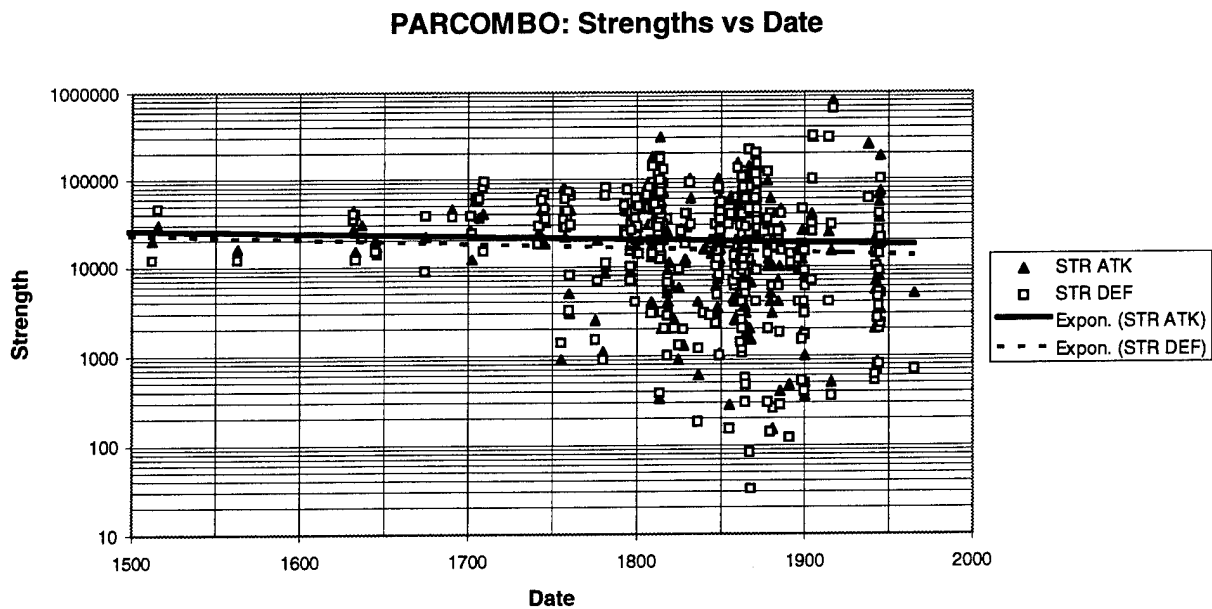
Figure 3-6. Attacker and Defender Strengths versus Date for the PARCOMBO Data Base

CDB91DAT: FRY vs Date
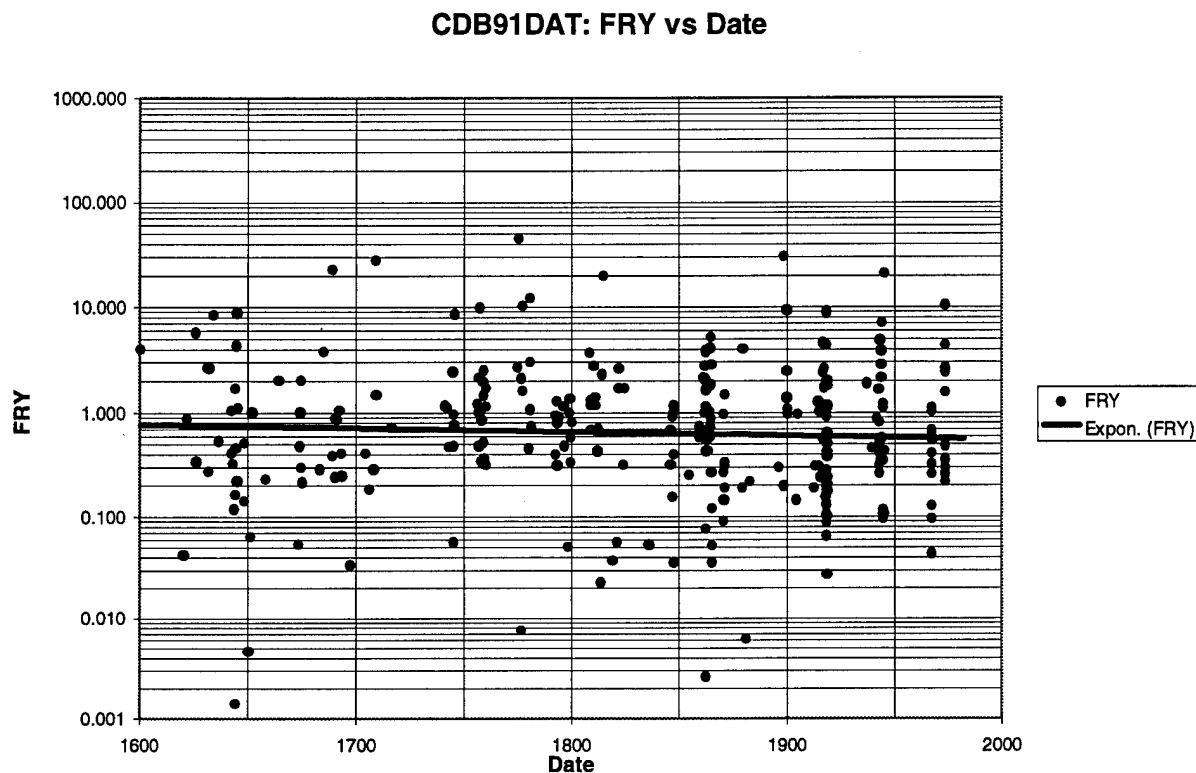
Figure 3-7. Force Ratio Favoring Defender versus Date for the CDB91DAT Data Base

PARCOMBO: FRY vs Date
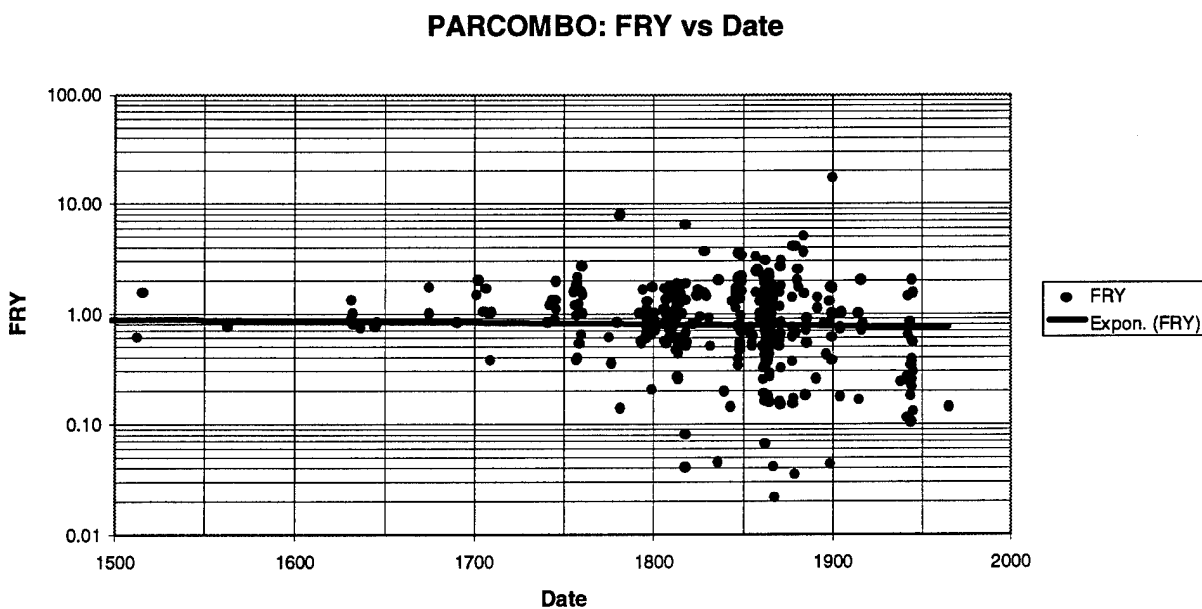
Figure 3-8. Force Ratio Favoring Defender versus Date for the PARCOMBO Data Base

CDB91DAT: TBC vs Date
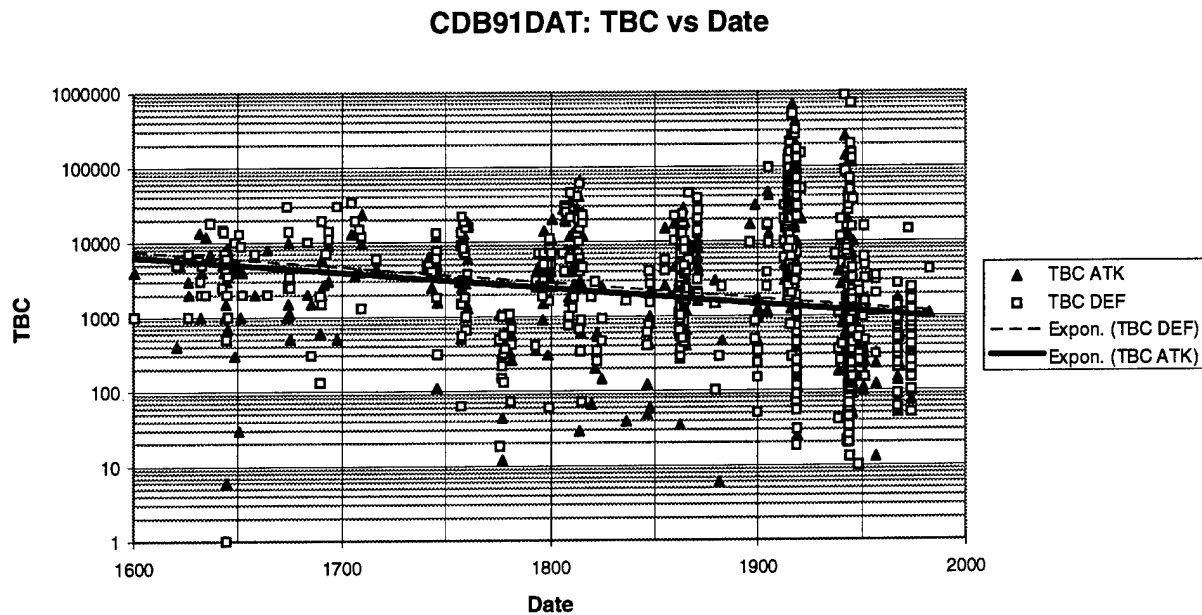
Figure 3-9. Attacker and Defender TBC versus Date for the CDB91DAT Data Base

PARCOMBO: TBC vs Date
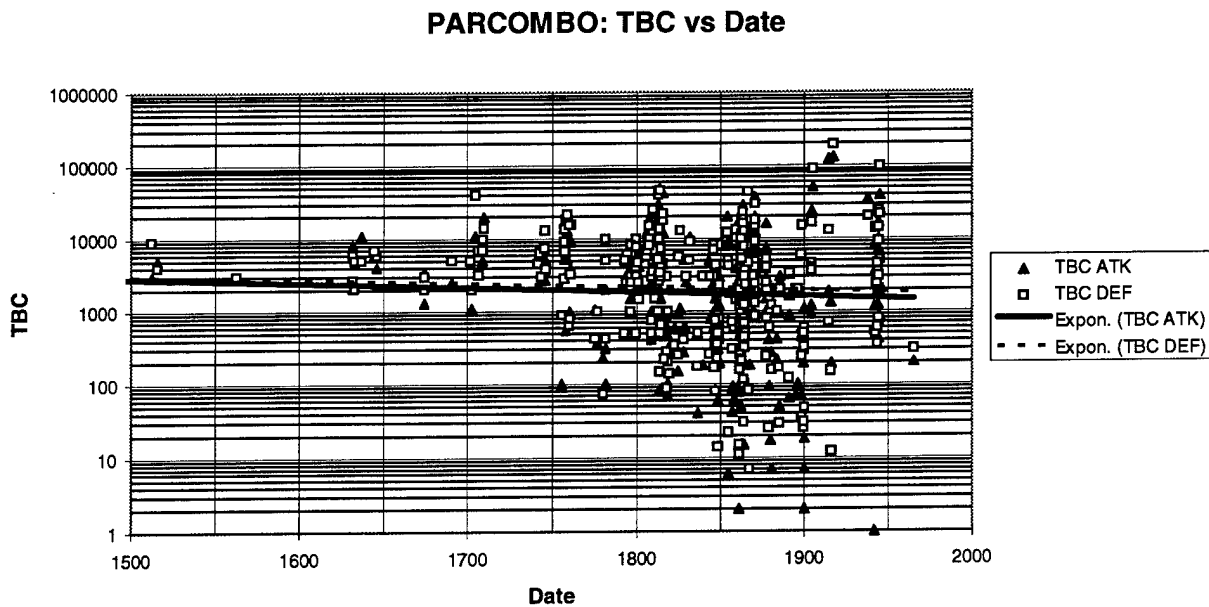
Figure 3-10. Attacker and Defender TBC versus Date for the PARCOMBO Data Base

CDB91DAT: CERY vs Date
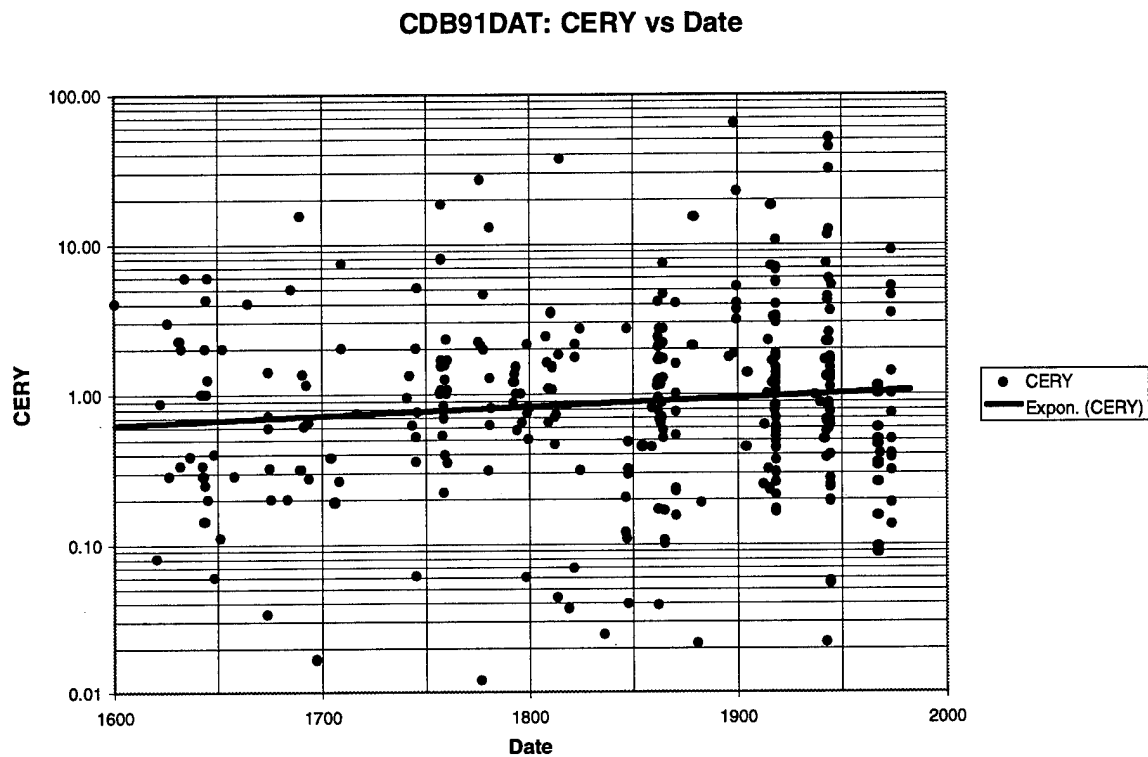
Figure 3-11. Casualty Exchange Ratio Favoring the Defender versus Date for the CDB91DAT Data Base

PARCOMBO: CERY vs Date
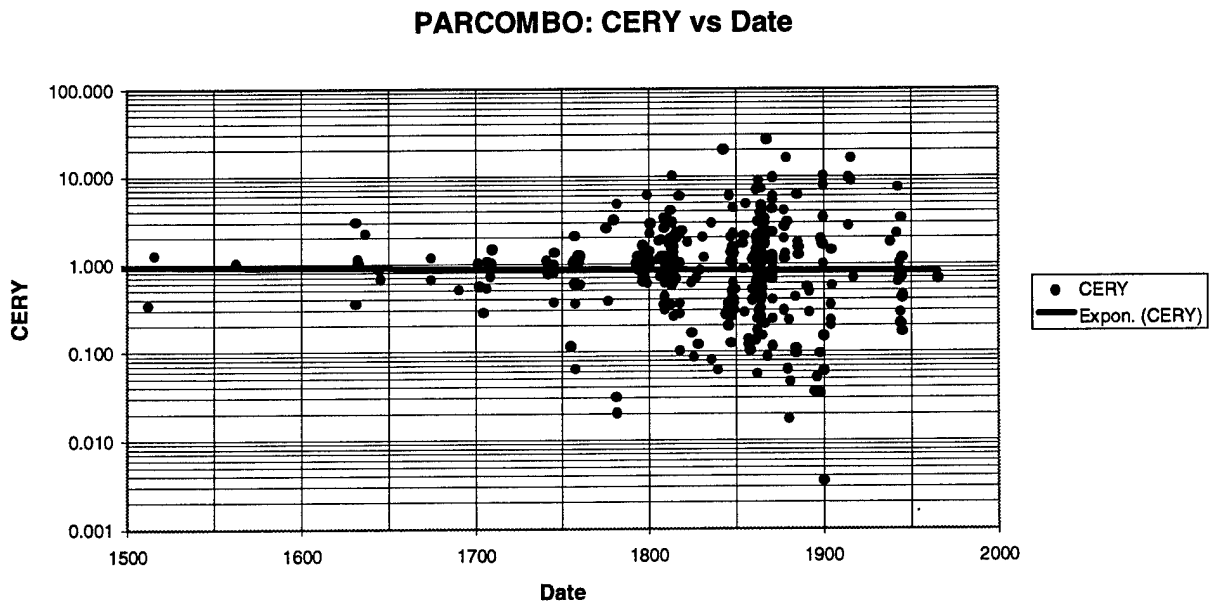
Figure 3-12. Casualty Exchange Ratio Favoring the Defender versus Date for the PARCOMBO Data Base

CDB91DAT: TBC Fractions vs Date
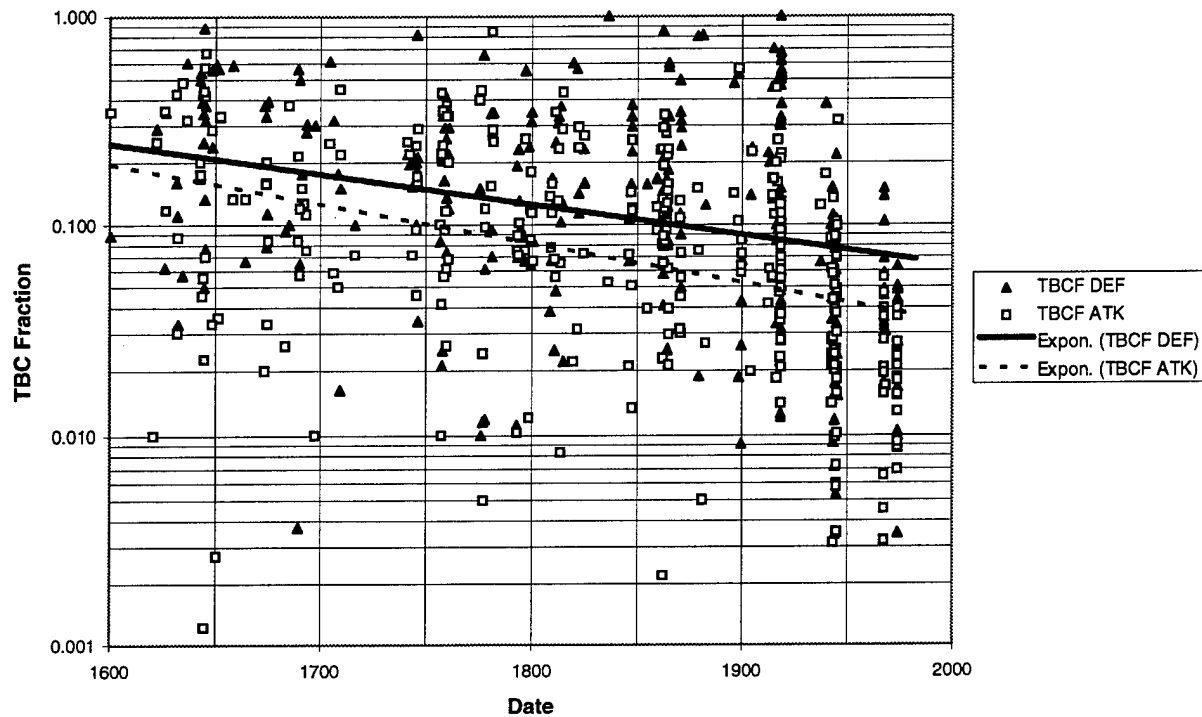
Figure 3-13. Attacker and Defender TBC Fractions versus date for the CDB91DAT
Data Base

PARCOMBO: TBC Fractions vs Date
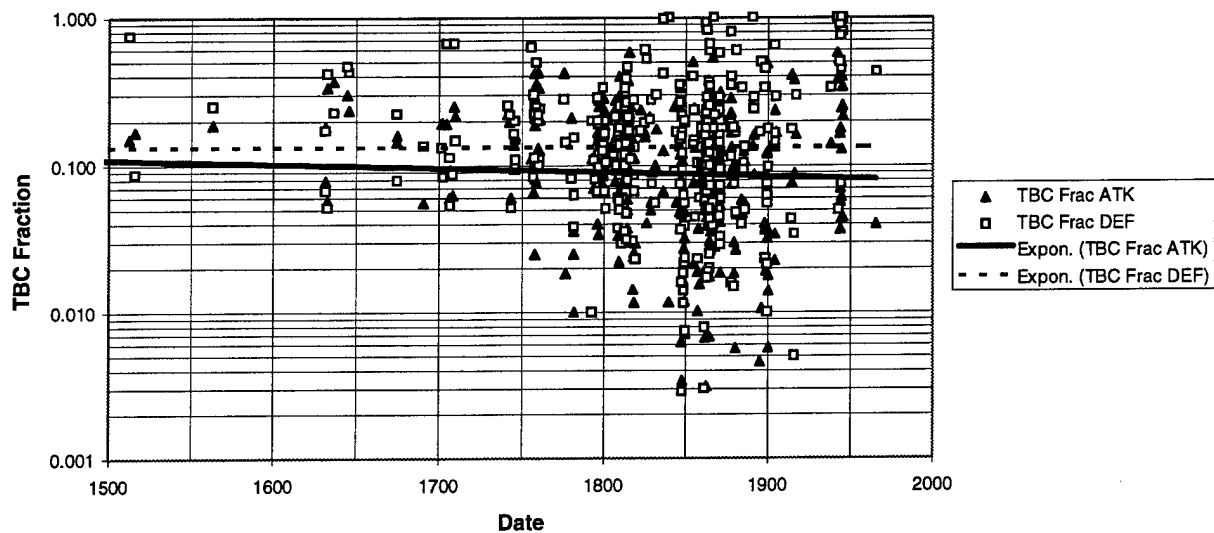
Figure 3-14. Attacker and Defender TBC Fractions versus Date for the PARCOMBO
Data Base

CDB91DAT: FERY vs Date
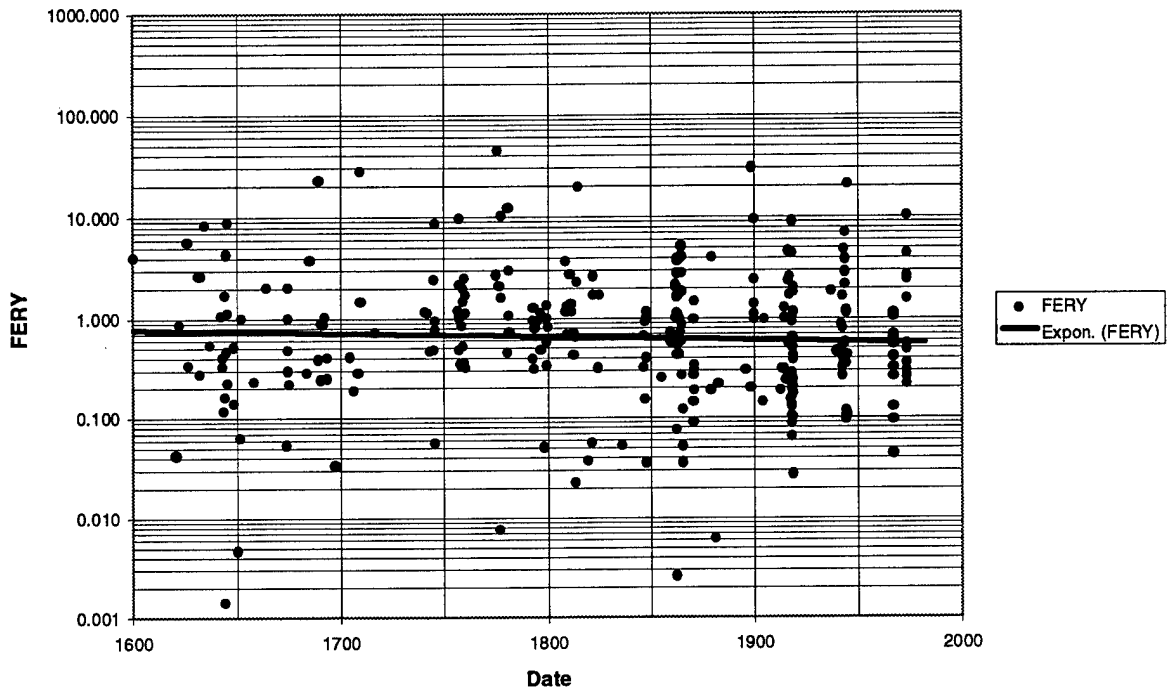
Figure 3-15. Fractional Exchange Ratio Favoring the Defender versus Date for the
CDB91DAT Data Base

PARCOMBO: FERY vs Date
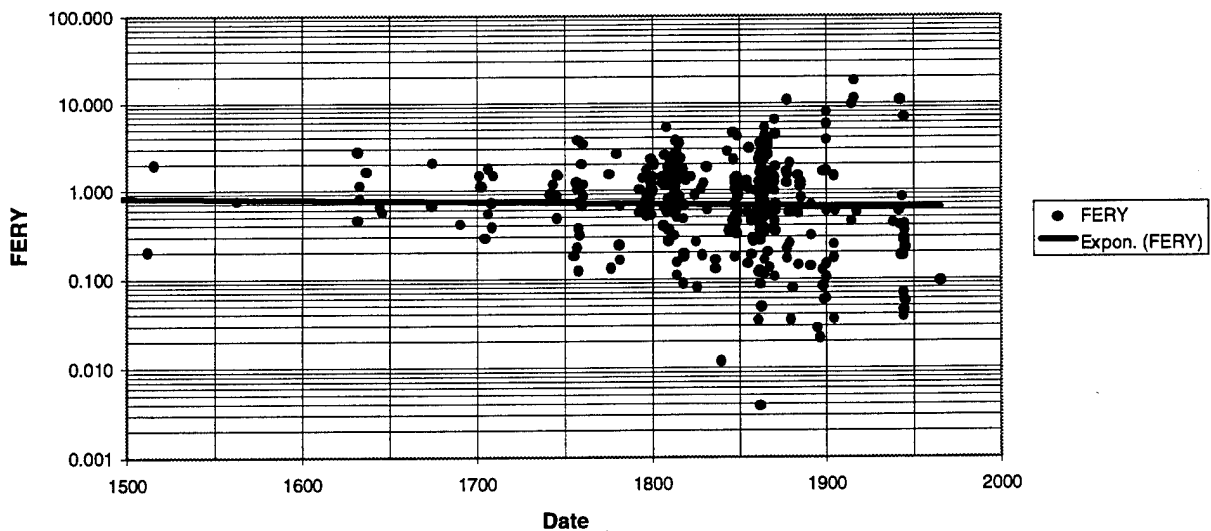
Figure 3-16. Fractional Exchange Ratio Favoring the Defender versus Date for the
PARCOMBO Data Base

CAA-RP-95-1

CDB91DAT: WIN/LOS Strengths vs Date
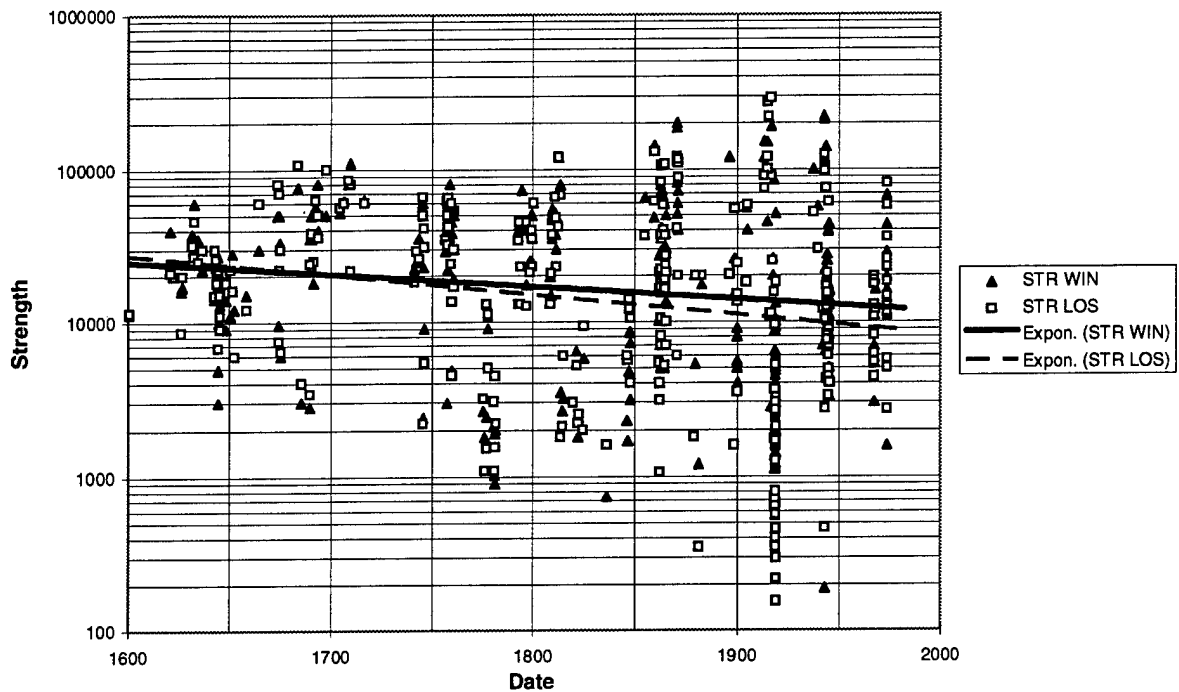
Figure 3-17. Winner and Loser Strengths versus Date for the CDB91DAT Data Base

PARCOMBO: WIN/LOS Strengths vs Date
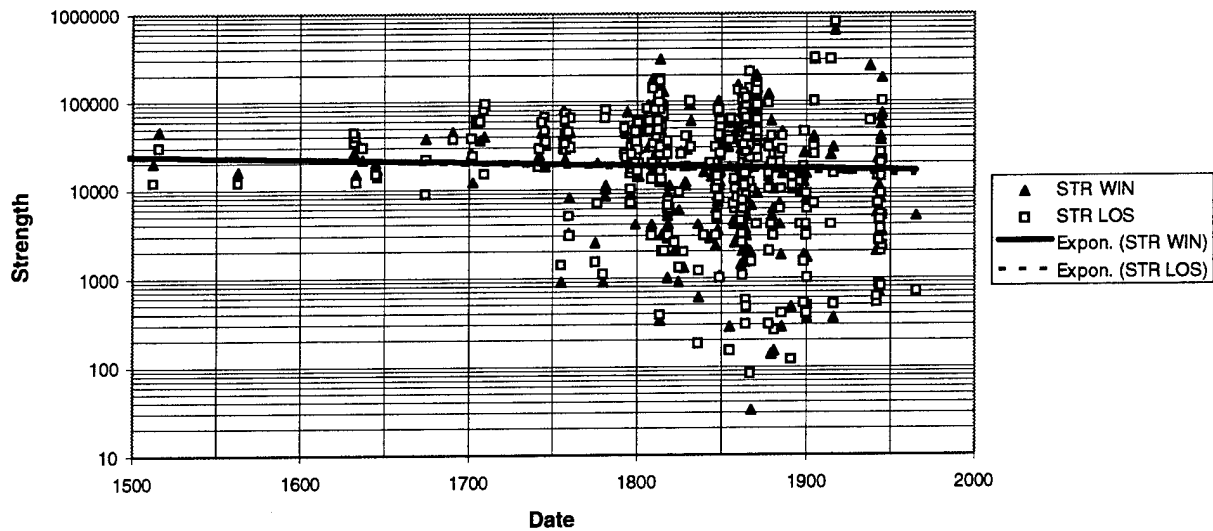
Figure 3-18. Winner and Loser Strengths versus Date for the PARCOMBO Data Base

3-16

BWSHALL: WIN/LOS Strengths vs Date
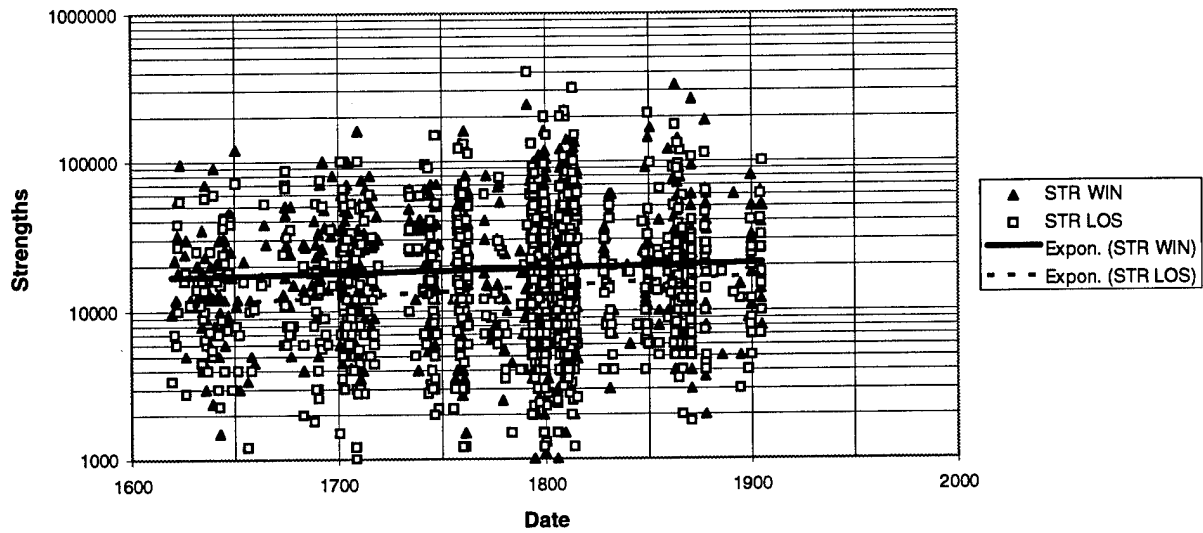
Figure 3-19. Winner and Loser Strengths versus Date for the BWSHALL Data Base

BODASHIP: WIN/LOS Strengths vs date
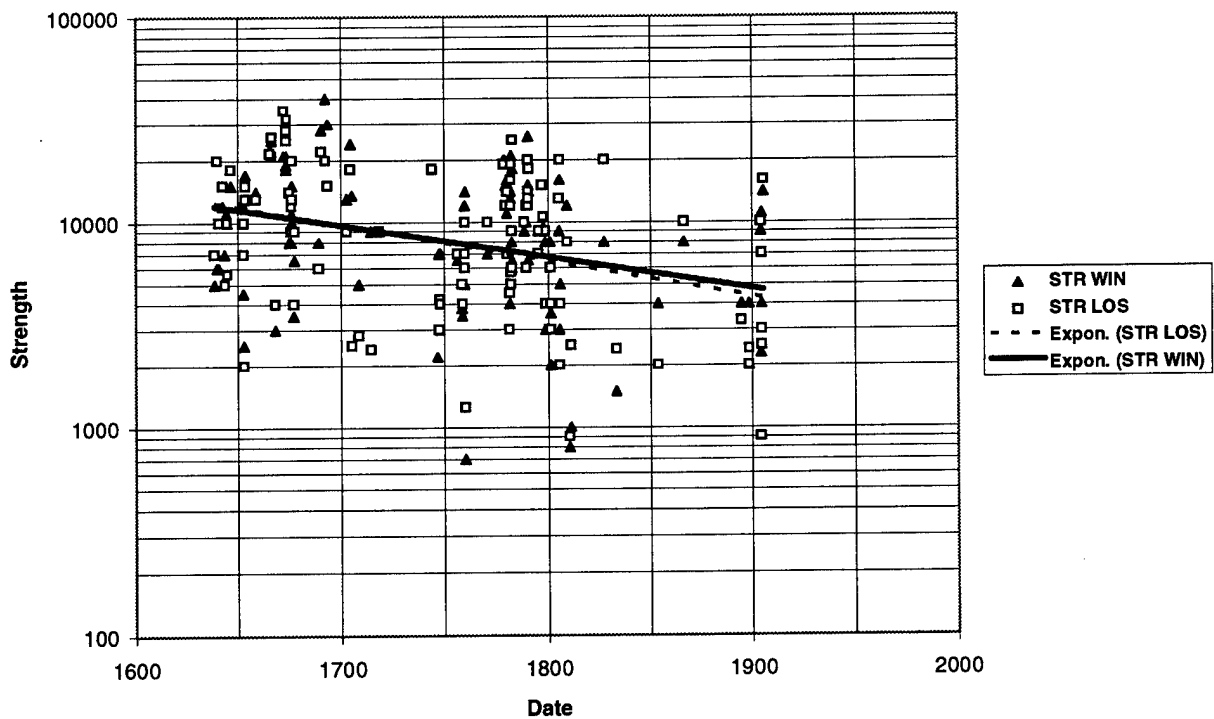
Figure 3-20. Winner and Loser Strengths versus Date for the BODASHIP Data Base

CDB91DAT: FRW vs Date
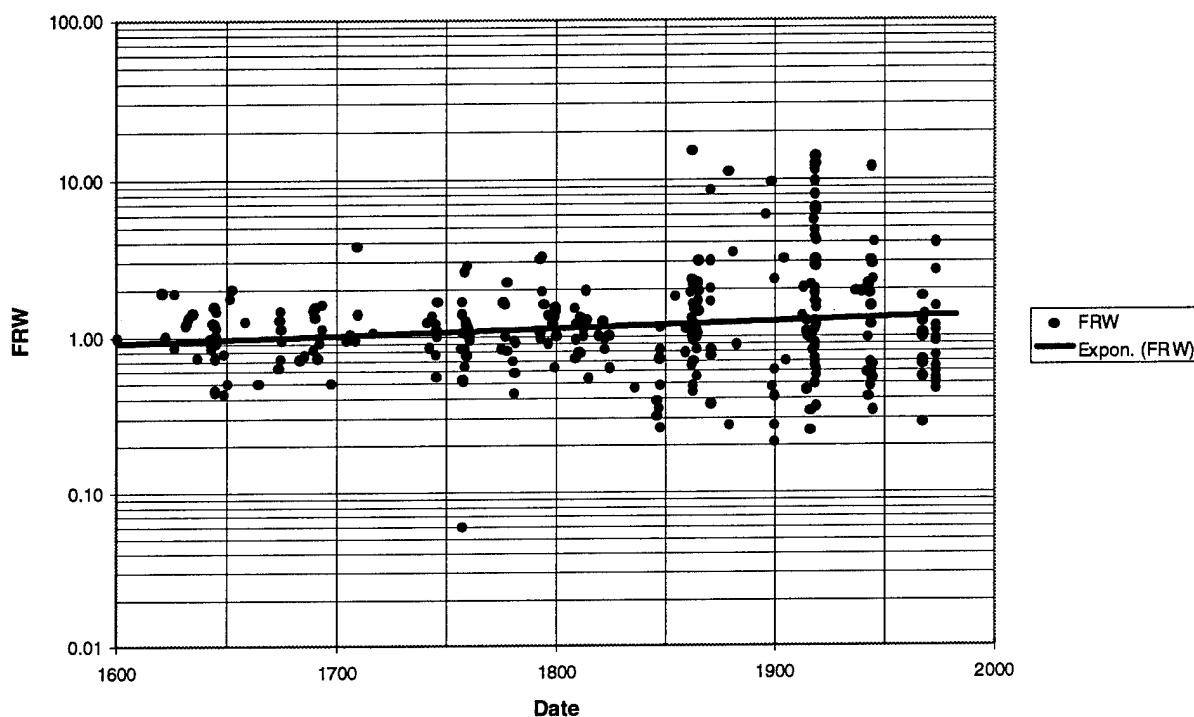
Figure 3-21. Force Ratio Favoring the Winner versus Date for the CDB91DAT Data Base

PARCOMBO: FRW vs Date
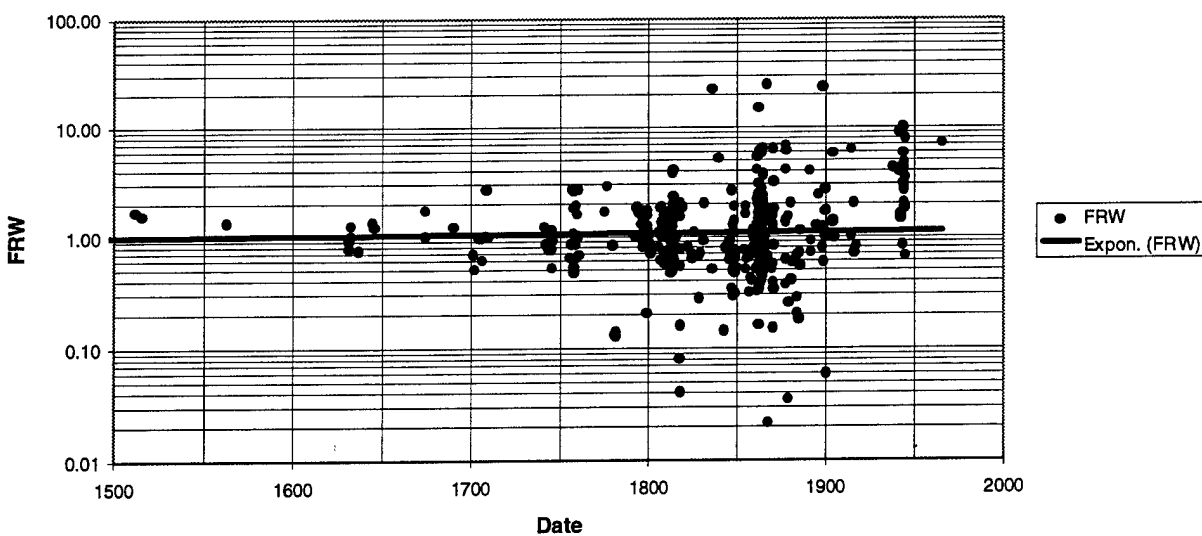
Figure 3-22. Force Ratio Favoring the Winner versus Date for the PARCOMBO Data Base

BWSHALL: FRW vs Date
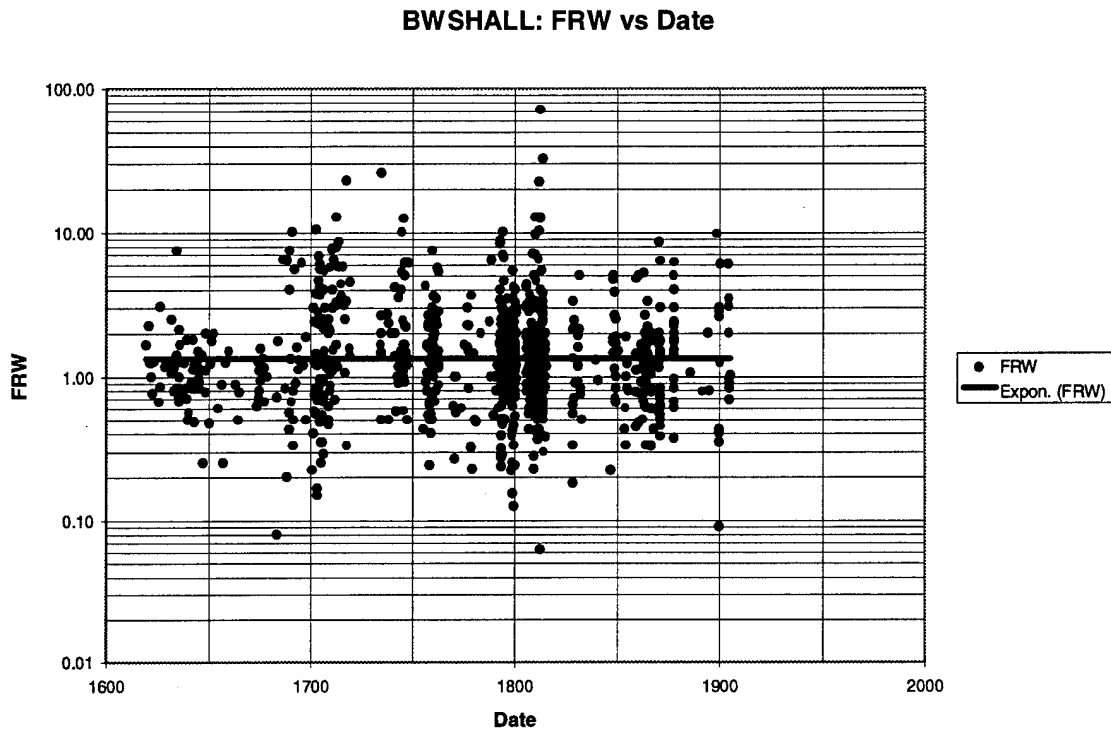
Figure 3-23. Force Ratio Favoring the Winner versus Date for the BWSHALL Data Base

BODASHIP: FRW vs Date
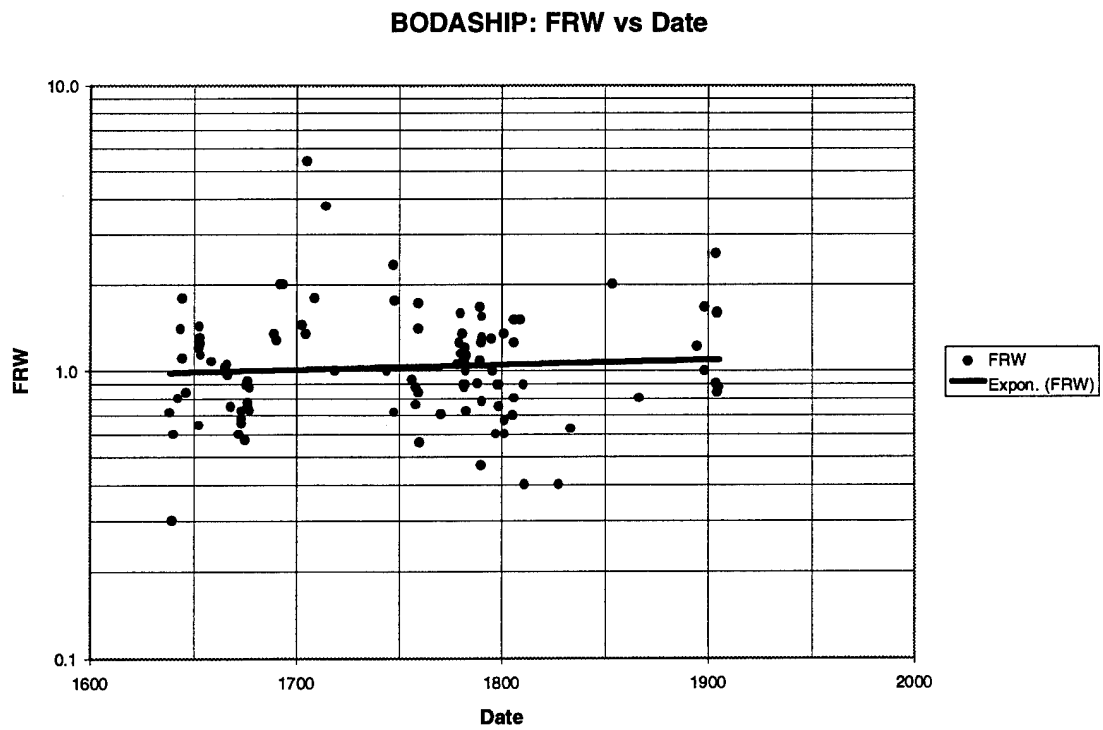
Figure 3-24. Force Ratio Favoring the Winner versus Date for the BODASHIP Data Base

CAA-RP-95-1

CDB91DAT: WIN/LOS TBC vs Date
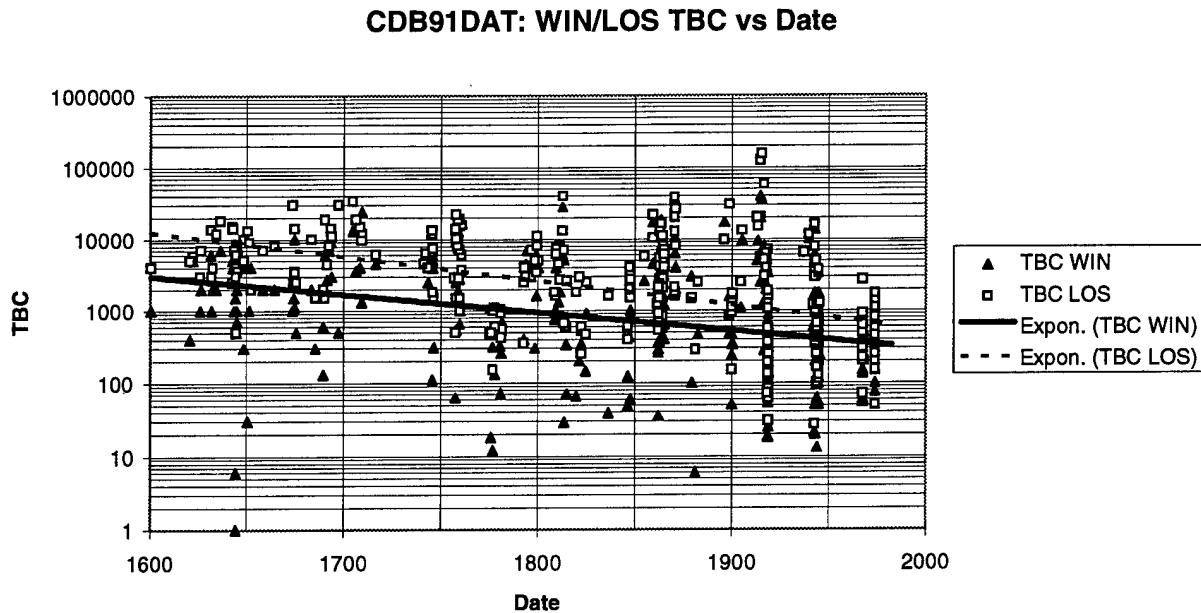
Figure 3-25. Winner and Loser TBC versus Date for the CDB91DAT Data Base

PARCOMBO: WIN/LOS TBC vs Date
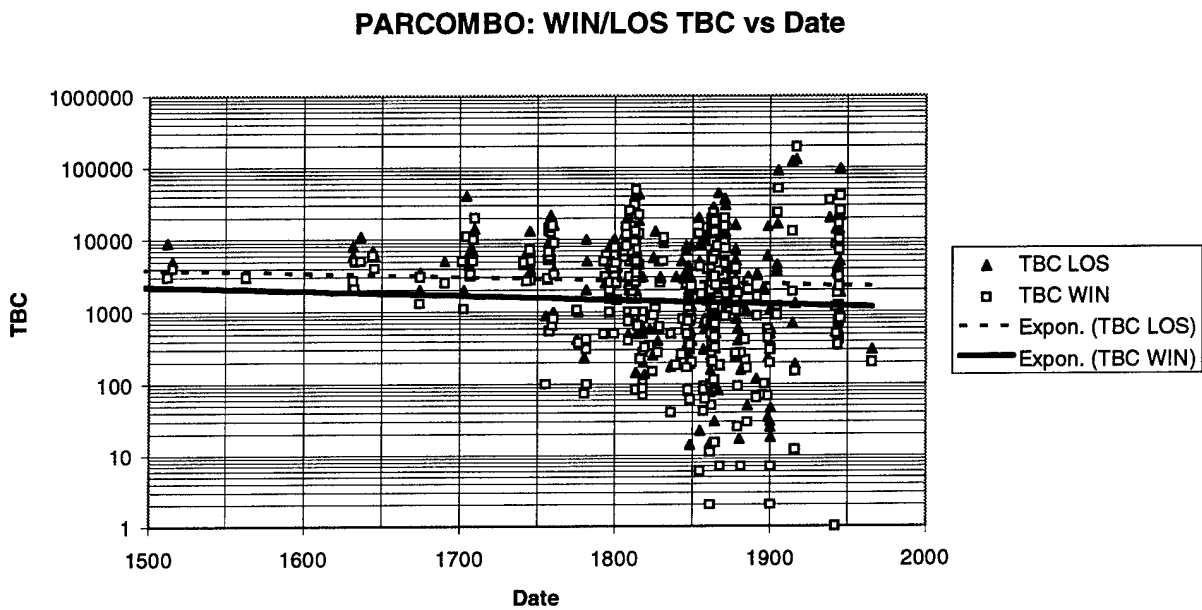
Figure 3-26. Winner and Loser TBC versus Date for the PARCOMBO Data Base

3-20

BWSHALL: WIN/LOS TBC vs Date
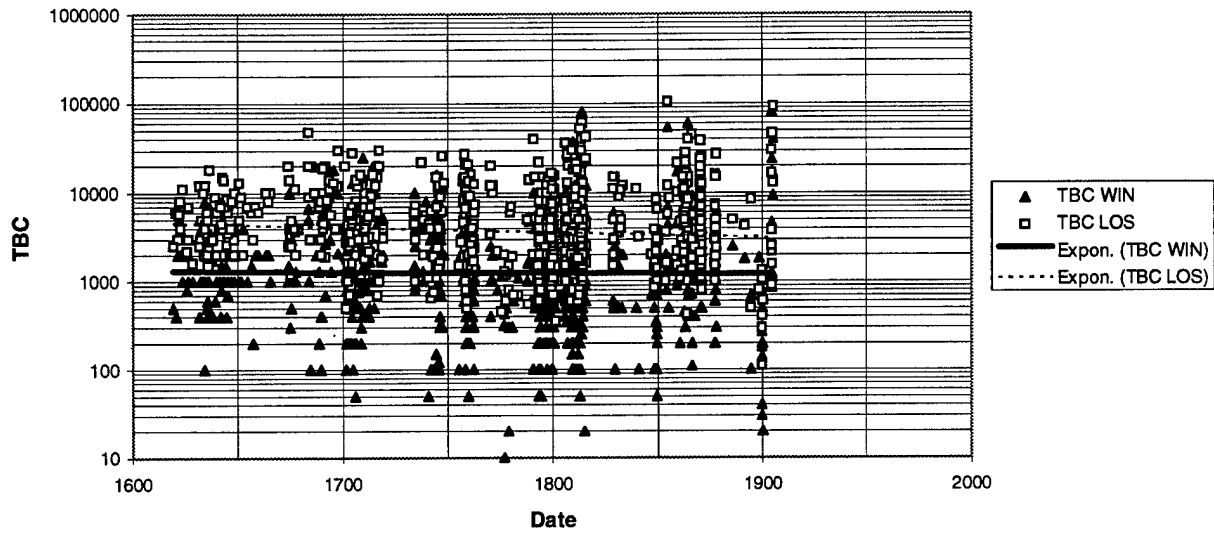
Figure 3-27. Winner and Loser TBC versus Date for the BWSHALL Data Base

BODASHIP: WIN/LOS TBC vs date
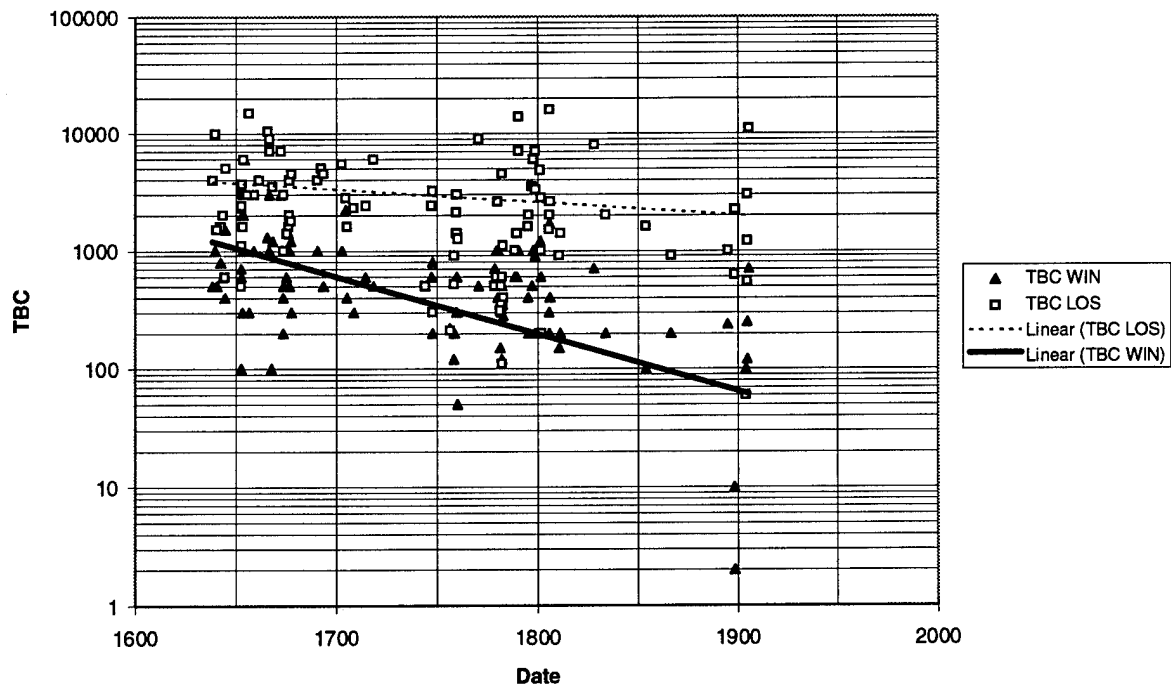
Figure 3-28. Winner and Loser TBC versus Date for the BODASHIP Data Base

CDB91DAT: CERW vs Date
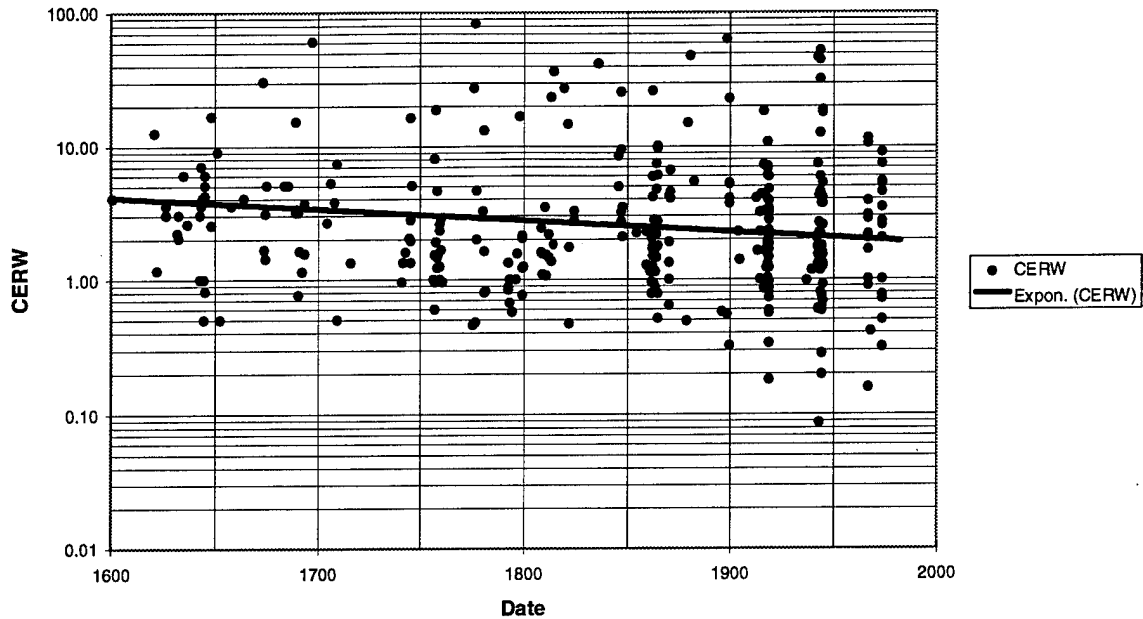
Figure 3-29. Casualty Exchange Ratio Favoring the Winner versus Date for the CDB91DAT Data Base

PARCOMBO: CERW vs Date
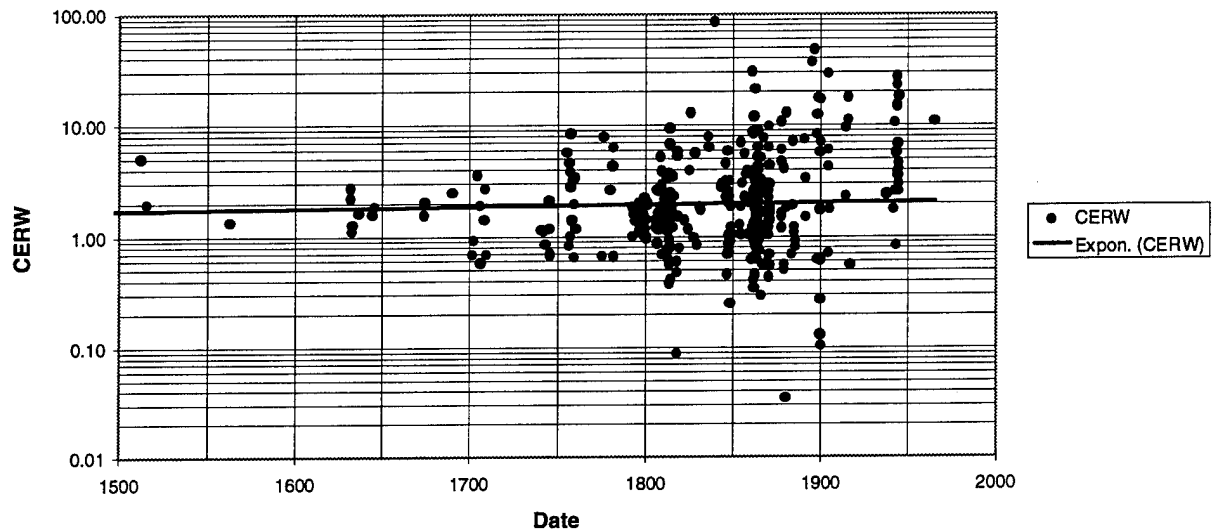
Figure 3-30. Casualty Exchange Ratio Favoring the Winner versus Date for the PARCOMBO Data Base

BWSHALL: CERW vs Date
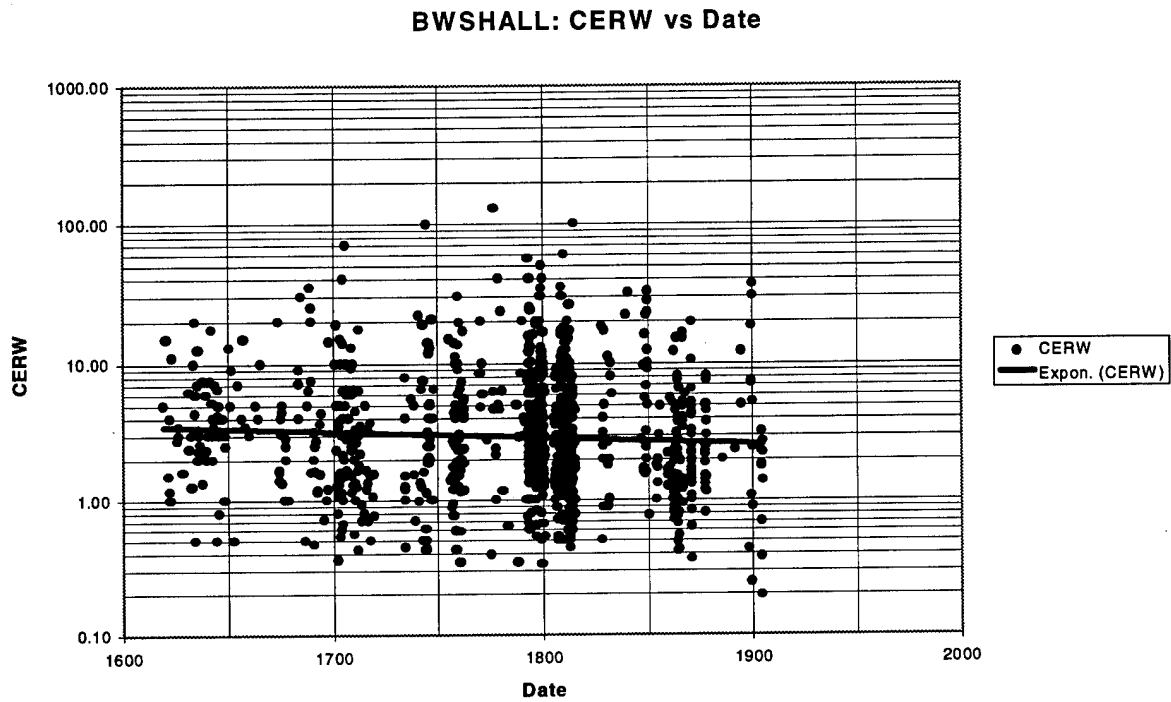
Figure 3-31. Casualty Exchange Ratio Favoring the Winner versus Date for the BWSHALL Data Base

BODASHIP: CERW vs Date
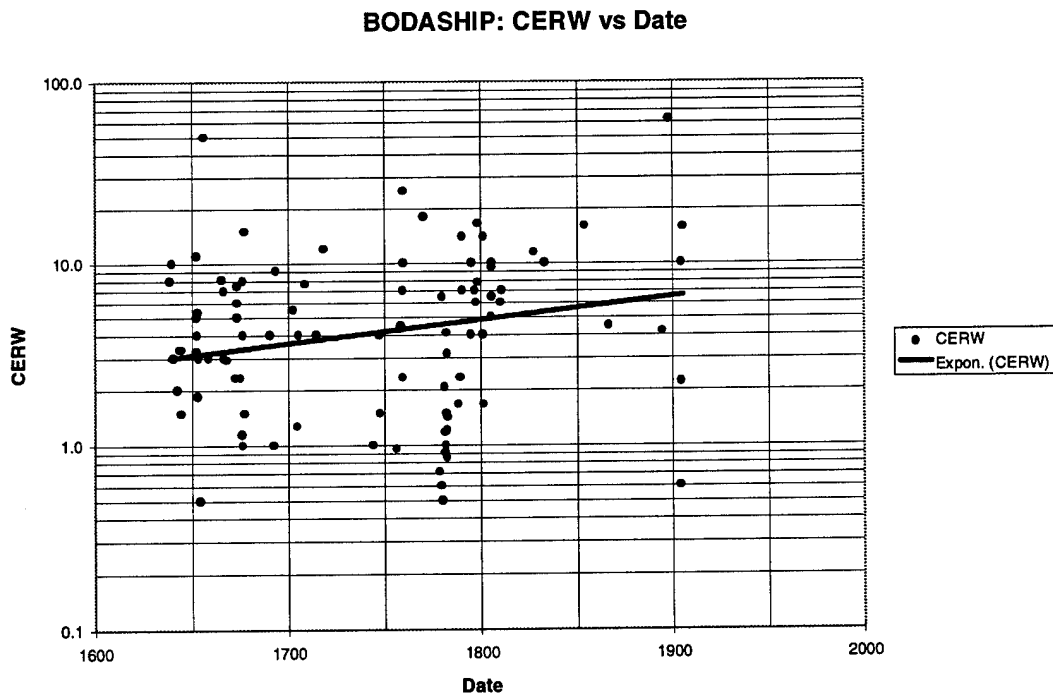
Figure 3-32. Casualty Exchange Ratio Favoring the Winner versus Date for the BODASHIP Data Base

CDB91DAT: WIN/LOS TBC Fractions vs Date
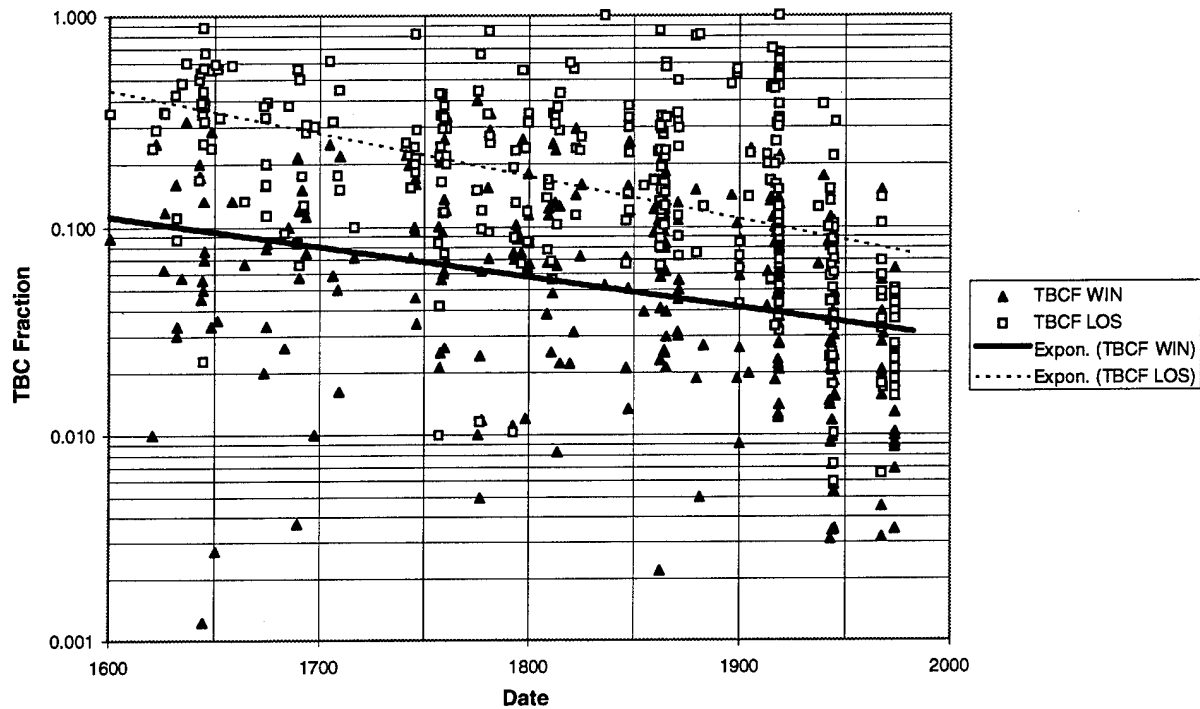
Figure 3-33. Winner and Loser TBC Fractions versus Date for the CDB91DAT Data Base

PARCOMBO: WIN/LOS TBC Fractions vs Date
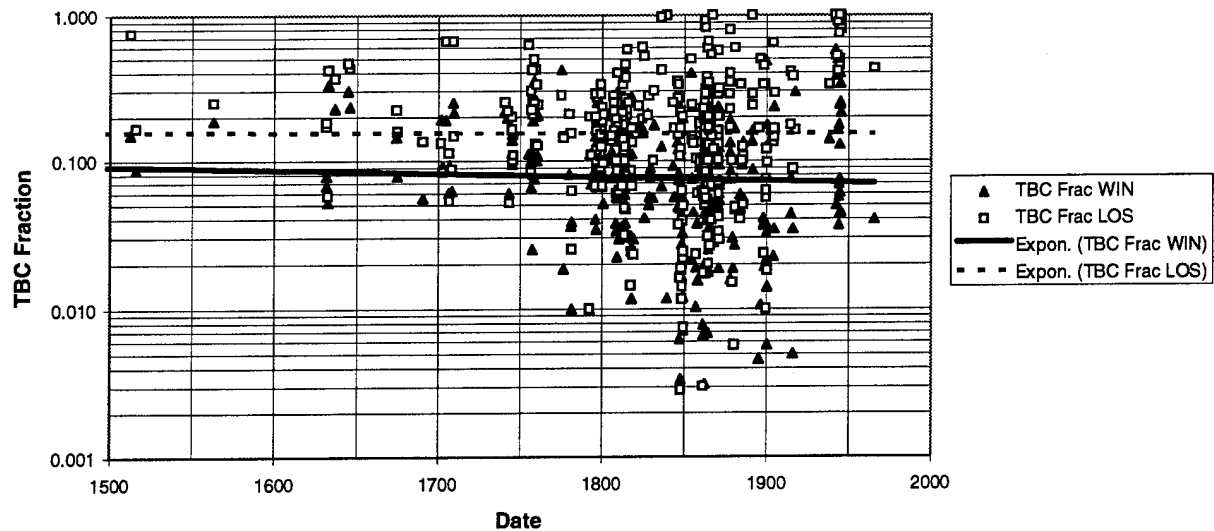
Figure 3-34. Winner and Loser TBC Fractions versus Date for the PARCOMBO Data Base

BWSHALL: WIN/LOS TBC Fractions vs Date

Figure 3-35. Winner and Loser TBC Fractions versus Date for the BWSHALL Data Base

BODASHIP: WIN/LOS TBC Fractions vs Date

Figure 3-36. Winner and Loser TBC Fractions versus Date for the BODASHIP Data Base
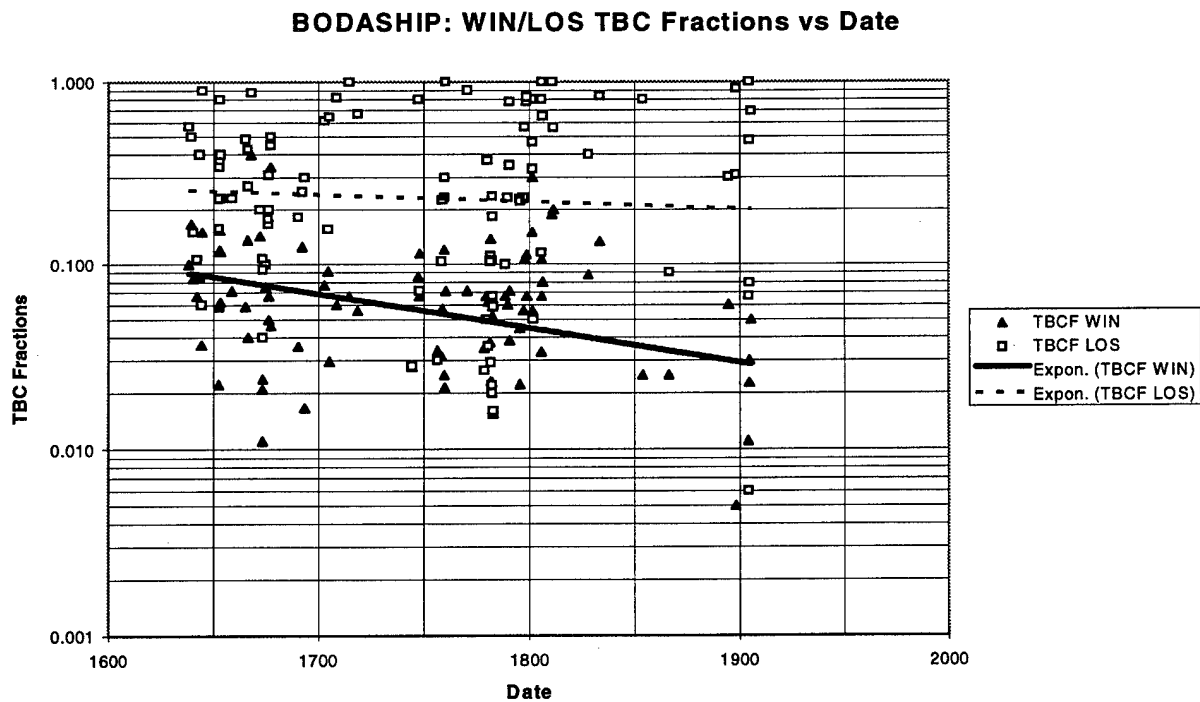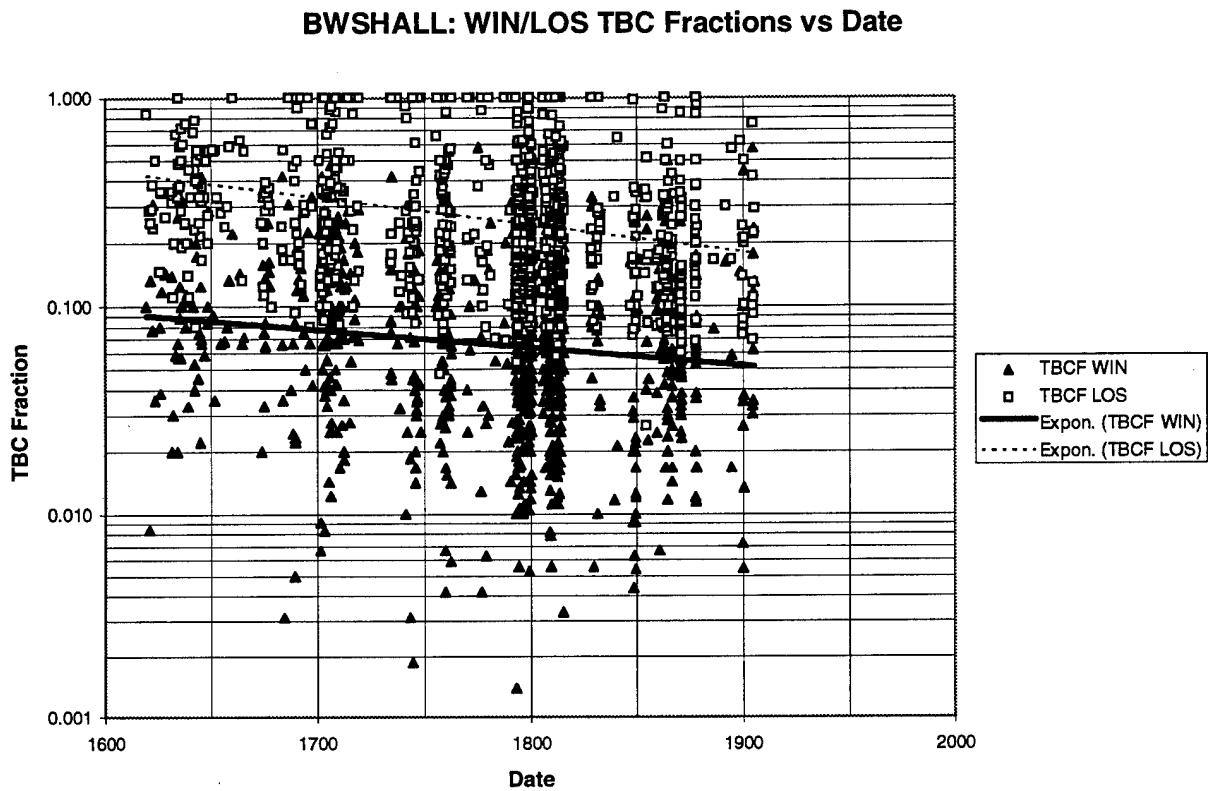
CDB91DAT: FERW vs Date
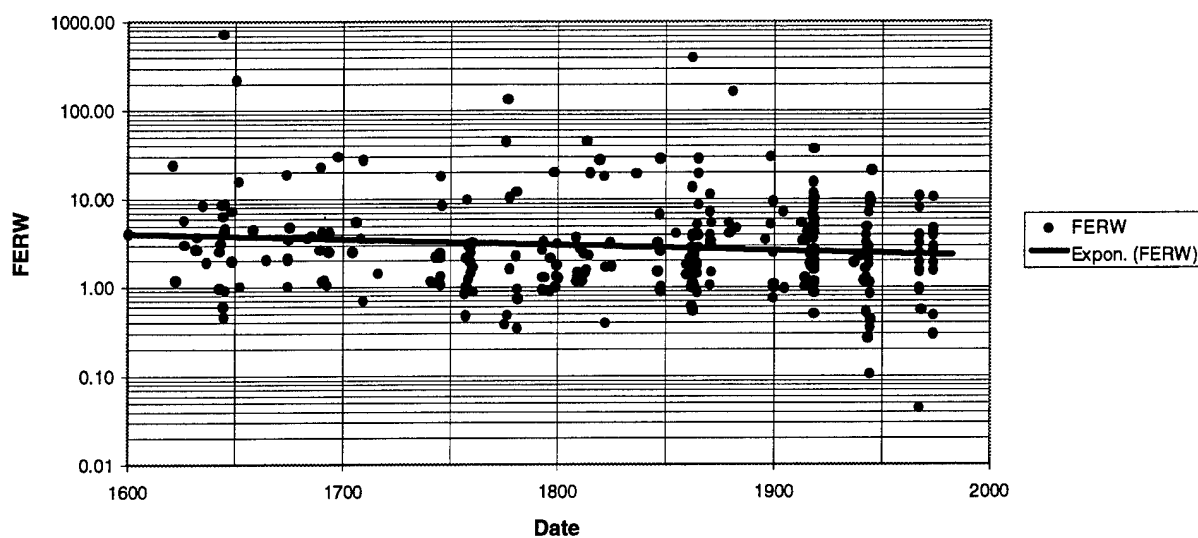
Figure 3-37. Fractional Exchange Ratio Favoring the Winner versus Date for the
CDB91DAT Data Base

PARCOMBO: FERW vs Date
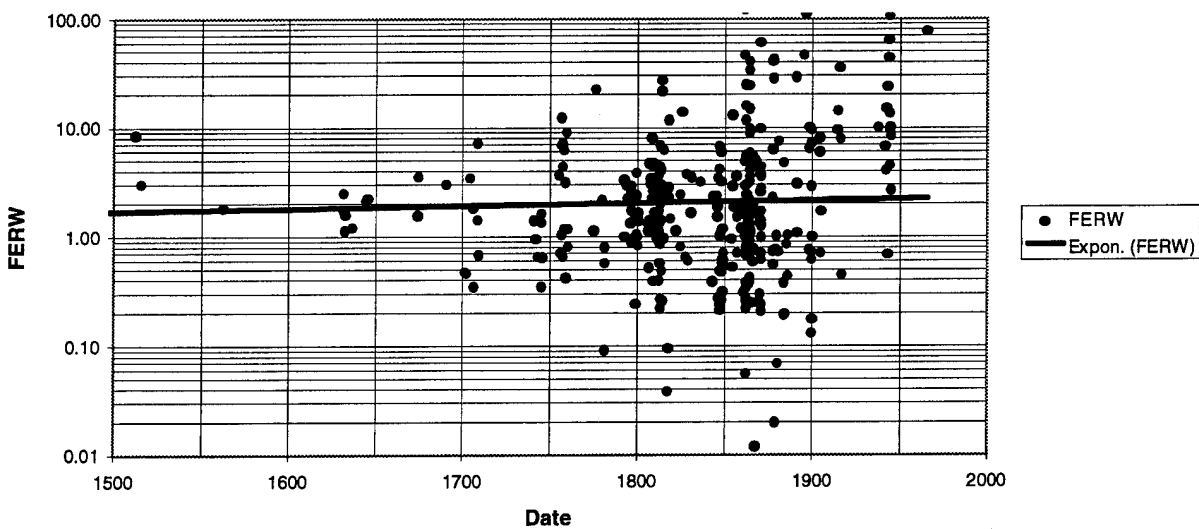
Figure 3-38. Fractional Exchange Ratio Favoring the Winner versus Date for the
PARCOMBO Data Base
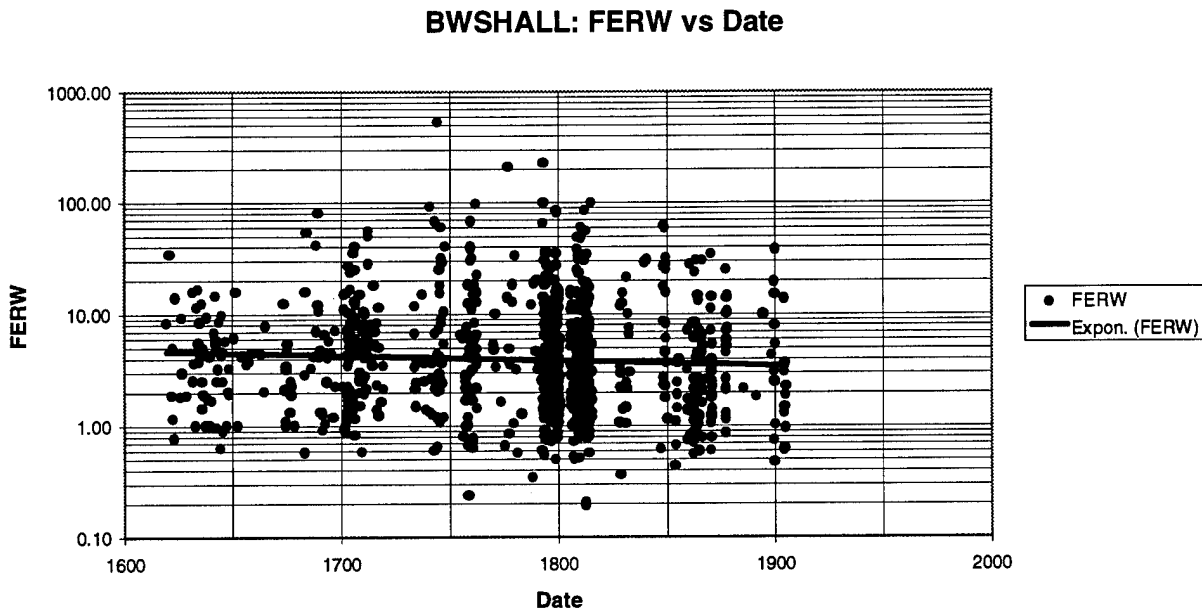
Figure 3-39. Fractional Exchange Ratio Favoring the Winner versus Date for the
BWSHALL Data Base
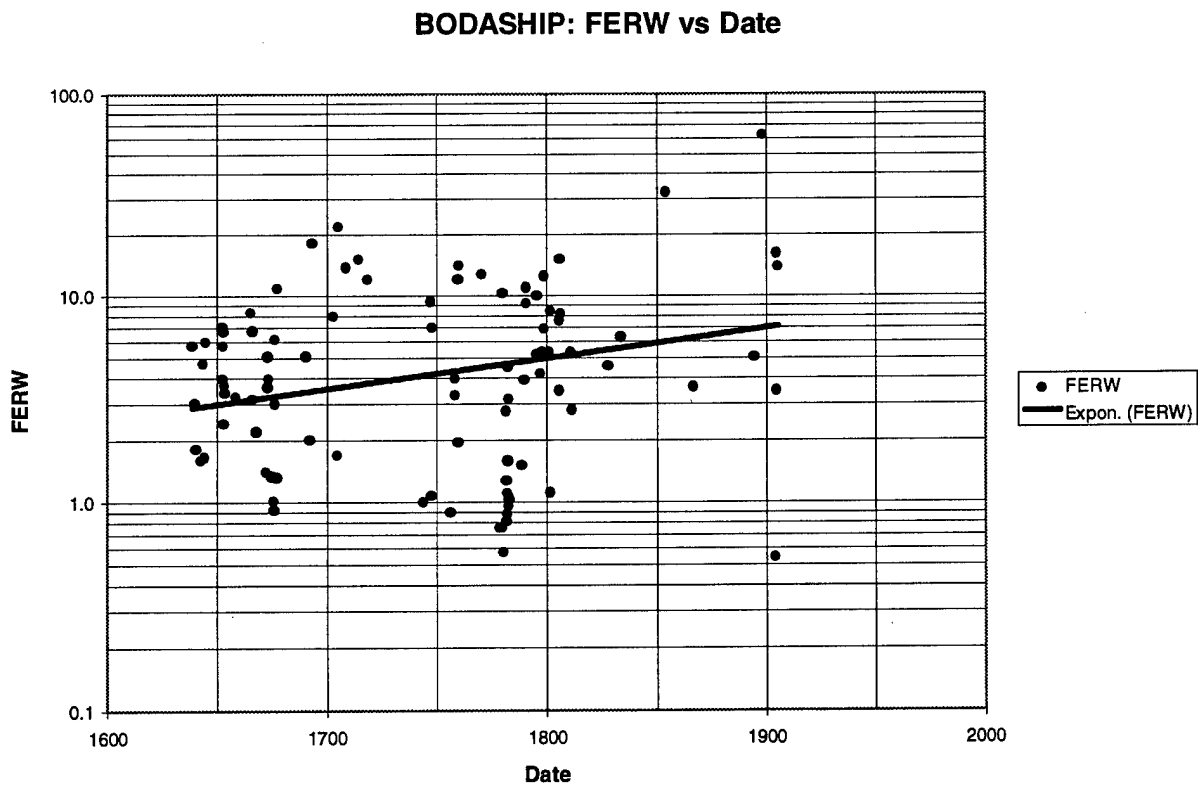
Figure 3-40. Fractional Exchange Ratio Favoring the Winner versus Date for the
BODASHIP Data Base

CHAPTER 4

ISSUE GROUP 2:
GENERAL RELATIONSHIPS AMONG
FORCE SIZES, BATTLE DURATIONS, AND CASUALTIES

4-1. INTRODUCTION. This chapter is devoted to the general question, "How are force sizes, battle durations, force ratios, casualty numbers, casualty fractions, and casualty rates of the opposing sides interrelated?"

4-2. NOTATIONAL CONVENTIONS. Later in this chapter we will briefly review the chief works on the relations of casualties to strengths to indicate our present understanding of this matter, but without attempting to provide a comprehensive review or critique of the current state of affairs. However, for the sake of clarity, we must convert their widely varied notations to a common denominator. Accordingly, we begin by introducing the notational conventions that will be used throughout this work.

    a. Notational Conventions for Strengths. The abbreviations ATK, DEF, WIN, and LOS will sometimes be used for the attacker, defender, winner, and loser, respectively. We use X to stand for the attacker's strength, Y to stand for the defender's, W the winner's, and L the loser's. Both upper and lower cases will be used interchangeably for the symbols denoting the sides— they are not case-sensitive. We also use Z to stand for the strength of an arbitrary side, and $\bar{Z}$ its opponent. Thus, by its definition,

$$\bar{Z} = \begin{cases} Y \text{ if } Z = X \\ X \text{ if } Z = Y \\ L \text{ if } Z = W \\ W \text{ if } Z = L \end{cases}$$

These symbols are used as postfixes, as in CX (the number of casualties to side X) and FX (side X's casualty fraction). A suffix or postfix may be used to modify these symbols. Thus, for example,

$$X0 = \text{ATK initial strength}$$
$$Y0 = \text{DEF initial strength}$$
$$W0 = \text{WIN initial strength}$$
$$L0 = \text{LOS initial strength}$$

Note that in other contexts strength may be the initial strength, the total strength, the average strength, and so forth, depending on the context. However, in this paper strengths are always considered to be initial personnel strengths, and so are measured in number of personnel. The force ratio favoring side Z is defined to be

CAA-RP-95-1

$$FRZ = \frac{Z0}{\overline{Z}0}.$$

We will almost always be concerned with the force ratio favoring the defender (side Y), that is, with

$$FRY = \frac{Y0}{X0}.$$

b. Notational Conventions for Casualties. We use C to represent a casualty number. Thus, for example, CZ stands for the number of casualties to side Z. Here casualties may be KIA, Bloody = KIA + WIA, TBC = Bloody + CMIA, or whatever, depending on the context. The casualty exchange ratio favoring side Z is defined to be

$$CERZ = \frac{C\overline{Z}}{CZ}.$$

In particular, the casualty exchange ratio favoring the defender (side Y) is

$$CERY = \frac{CX}{CY}.$$

Side Z's casualty fraction is defined to be

$$FZ = \frac{CZ}{Z0}.$$

The fractional exchange ratio favoring side Z is defined to be the ratio of the casualty fractions favoring side Z that is,

$$FERZ \equiv \frac{F\overline{Z}}{FZ} = CERZ \times FRZ.$$

In particular, the fractional exchange ratio favoring the defender (side Y) is

$$FERY \equiv \frac{FX}{FY} \equiv \frac{CX}{CY}\frac{Y0}{X0} = CERY \times FRY.$$

c. Notations for Constants. The symbols K, Q, and P are used as generic constants. Thus, the statement that $x = Ky$ expresses the notion that x is proportional to y, with K being the constant of proportionality. Postfixes are used to identify individual constants when more than one must be considered. For example, KYX may be used as the coefficient of $\ln X0$ in an expression for $\ln CY$.

d. Notations for Regression Models Relating Casualties to Strengths, Battle Durations, and Battle Dates. For the purposes of this paper, we adopt the following basic regression model relating casualties to strengths, durations, and dates:

$$\ln CZ = KZ0 + KZZ \cdot \ln Z0 + KZ\overline{Z} \cdot \ln \overline{Z}0 + KZT \cdot T + KZD \cdot D$$

where T is the battle duration in days (duration may be overall, elapsed, active, or whatever, depending on the context), and D is the centered and scaled battle date given by

$$D = (BattleDate - 1800) / 100,$$

where *BattleDate* is the battle date (in years Anno Domini and fractions thereof). We adopt 1800 AD as the standard "zero year" and the century (100 years) as the basic date unit. This basic model corresponds to the relation

$$CZ = e^{KZ0 + KZT \cdot T + KZD \cdot D} \cdot (Z0)^{KZZ} \cdot (\overline{Z}0)^{KZ\overline{Z}}.$$

Thus, e^{KZ0} corresponds to the (theoretical) number of casualties to side Z when $T = 0$ days, $D = 0$ so that *BattleDate* = 1800, and $Z0 = \overline{Z}0 = 1$ (person). Specifically, the regression model reads

$$\ln CX = KX0 + KXX \cdot \ln X0 + KXY \cdot \ln Y0 + KXT \cdot T + KXD \cdot D$$

$$\ln CY = KY0 + KYX \cdot \ln X0 + KYY \cdot \ln Y0 + KYT \cdot T + KYD \cdot D$$

$$\ln CW = KW0 + KWW \cdot \ln W0 + KWL \cdot \ln L0 + KWT \cdot T + KWD \cdot D$$

$$\ln CL = KL0 + KLW \cdot \ln W0 + KLL \cdot \ln L0 + KLT \cdot T + KLD \cdot D$$

e. Notation for Regression Models Relating Casualty and Fractional Exchange Ratios to Strengths, Battle Durations, and Battle Dates. The regression model adopted for the casualty exchange ratio favoring side Z,

$$CERZ = C\overline{Z} / CZ,$$

can be written as

$$\ln CERZ = QZ0 + QZZ \cdot \ln Z0 + QZ\overline{Z} \cdot \ln \overline{Z}0 + QZT \cdot T + QZD \cdot D,$$

where the Q's may in selected cases be expressible in terms of the K's in the preceding paragraphs as follows:

$$QZZ' = K\overline{Z}Z' - KZZ',$$

and it is understood that Z is usually taken to be either Y or W, while Z' runs through all of the values 0, $\overline{Z}$, Z, T, and D. For example, we have

$$\ln CERY \equiv \ln(CX / CY) = QY0 + QYY \cdot \ln Y0 + QYX \cdot \ln X0 + QYT \cdot T + QYD \cdot D,$$

where in some selected cases the Q's may be related to the K's via:

$$QY0 = KX0 - KY0$$
$$QYY = KXY - KYY$$
$$QYX = KXX - KYX$$
$$QYT = KXT - KYT$$
$$QYD = KXD - KYD$$

The relation connecting the Q's and the K's may be different from those given above in some cases. The reasons for that are explained in Appendix D.

Similar regression models are adopted for $\ln CERW$. The specific basic regression model equations for $CERY$ and $CERW$ are:

$$CERY = e^{QY0+QYT \cdot T+QYD \cdot D} \cdot (Y0)^{QYY} \cdot (X0)^{QYX}$$
$$CERW = e^{QW0+QWT \cdot T+QWD \cdot D} \cdot (W0)^{QWW} \cdot (L0)^{QWL}$$

4-3. ON THE RELATION OF REGRESSION MODELS TO DIFFERENTIAL EQUATION MODELS. There is much confusion in the literature abut the connections between regression models and differential equation models. The following remarks may help to clarify the situation.

a. To begin with, *there is no automatic connection between the constants in regression models and the constants in differential equation models.* To think or to behave otherwise is to commit the Constant Fallacy (see Helmbold-1994). Thus, a particular differential equation formulation does not necessarily imply a specific regression model, nor does a particular regression model necessarily imply a specific differential equation formulation. In particular, a differential equation formulation such as Lanchester's square law,

$$dx = -By\,dt$$
$$dy = -Ax\,dt$$

does not necessarily imply that the regression model is of the form

$$CX = B \cdot Y0 \cdot T$$
$$CY = A \cdot X0 \cdot T$$

Nor does a regression model of that form necessarily imply that Lanchester's square law differential equation formulation holds.

b. However, much past work has looked to a variety of differential equation formulations to inspire and suggest regression models. In fact, the motivation for much prior work cannot be understood without reference to the differential equation formulations that inspired it. Accordingly, we note such inspiration wherever appropriate. Indeed, we may ourselves make use of this device to suggest regression models. Nevertheless, it must be constantly borne in mind that these are merely loose analogical connections, having no logical validity and carrying no logical necessity.

4-4. SUMMARY OF PRIOR WORK. Our summary of prior work on the relation of casualties to strengths, durations, and battle dates is highly focused. We present only the principal results. Even within that narrow focus, we seek only to capture the one or two points of greatest relevance to the present paper. Accordingly, the reader should not be misled into thinking that the papers reviewed contain *only* the information we cite. On the contrary, each contains many important observations and insights worthy of attentive reading and thoughtful consideration, but which happen to be peripheral to our immediate interests. Table 4-2, presented later, summarizes some key values drawn from the following reviews of prior work.

a. On Defender's Casualties. There is an argument to the effect that the defender's casualties should be lower than the attacker's because the defender has certain inherent advantages. For example, Clausewitz-1976, pp 390-392, *passim* observes that (emphasis in the original) "The defender waits for the attack in position, having chosen a suitable area and prepared it; which means he has carefully reconnoitered it, erected solid defenses at some of the most important points, established and opened communications, sited his batteries, fortified some villages, selected covered assembly areas, and so forth. The strength of his front, access to which is barred by one or more parallel trenches or other obstacles or by dominant strong points, makes it possible for him, while the forces at the points of actual contact are destroying each other, *to inflict heavy losses on the enemy at low cost to himself* as the attack passes through the successive stages of resistance until it reaches the heart of the position. ... We maintain unequivocally that the form of warfare that we call defense not only offers greater probability of victory than attack, but that its victories can attain the same proportions and results. Moreover, this applies not only to the *aggregate success* of all engagements that make up a campaign, but to *each individual* battle, provided there is no lack of strength and determination." There seem to be two interpretations of this claim, depending on whether the "losses" referred to are casualty numbers or casualty fractions. The casualty number version is that CX is greater than CY, *i.e.,* that $CERY = CX/CY = K$, where K is a constant whose value is definitely greater than unity. The casualty fraction version is that $FERY = FX/FY = K$, where K is a constant whose value is definitely greater than unity. The casualty fraction version is mathematically equivalent to the formula $CERY = FERY \div FRY = K(FRY)^{-1}$, where K is a constant whose value is greater than unity and FRY stands for the initial force ratio favoring the defender, $Y0/X0$.

b. The Effect of Force Size on Casualties. Dodge-1900 suggests that losses to smaller units are disproportionately high, compared to larger units. Berndt-1897 and Bodart-1908 also mention this phenomenon, and it has since been popularized by Dupuy-1990. In mathematical terms, this is the suggestion that FX and FY are decreasing functions of $X0$ and $Y0$, respectively. Mathematically, this might be approximated by the equations $FX = KX \cdot X0^{PX}$ and

$FY = KY \cdot Y0^{PY}$, where the K's and P's are constants to be determined by the data, and the assertion is that both of the exponents (the P's) are less than unity—though, presumably, greater than zero. One implication of this is that $FERY \equiv \dfrac{FX}{FY} = K\dfrac{X0^{PX}}{Y0^{PY}}$, or that

$$CERY \equiv FERY \div FRY = K\frac{X0^{1+PX}}{Y0^{1+PY}} = K(FRY)^{-(1+PY)}X0^{PX-PY}.$$

c. Osipov's Law. Osipov-1915 uses historical data on 38 of the most notable battles of the 1800s and very early 1900s. This appears to be the first systematic scientific analysis of the relation of casualties to force sizes—that is, the first to use explicit mathematical models and statistical methods of data analysis. Osipov first derives an attrition model of the form

$$dx = -By\, dt$$
$$dy = -Ax\, dt$$

These equations, which later became known as Lanchester's Square Law, suggested to Osipov that the casualties are related by the regression model

$$CX \approx B \cdot Y0 \cdot T$$
$$CY \approx A \cdot X0 \cdot T$$

so that the casualty exchange ratio favoring side Y would be related to the force ratio favoring side Y by the regression model $CERY \approx K \cdot FRY$. Osipov notes that this regression model is in relatively poor agreement with the historical data. Indeed, he shows that a much better fit is provided by the regression model suggested by an attrition model of the form

$$dx = -B\sqrt{y}\, dt$$
$$dy = -A\sqrt{x}\, dt$$

which corresponds to the regression model

$$CX \approx B\sqrt{Y0} \cdot T$$
$$CY \approx A\sqrt{X0} \cdot T$$

so that the casualty exchange ratio favoring side Y would be related to the force ratio favoring side Y by the regression model $CERY \approx K(FRY)^{1/2}$. These differential equations and regression models are found specifically in Osipov's Comments 1 and 3, made in connection with his formulas (12) and (6-bis). They have been called "Osipov's Law," by analogy to Lanchester's Law.

Thus, Osipov's Law asserts that the regression coefficients in

$$\ln CERY \equiv \ln(CX / CY) = QY0 + QYY \cdot \ln Y0 + QYX \cdot \ln X0,$$

are

$$QYY = 1/2$$
$$QYX \equiv -QYY = -1/2$$

Osipov makes no specific comment about the value of $QY0$.

d. Lanchester's Laws. Lanchester's Square Law appeared about the same time as Osipov's (see Lanchester-1916). It suggests that the coefficients in

$$\ln CERY \equiv \ln(CX / CY) = QY0 + QYY \cdot \ln Y0 + QYX \cdot \ln X0$$

are

$$QYY = +1$$
$$QYX = -1$$

In addition, Lanchester proposed the following Linear Law differential equation formulation

$$dx = -Byx\, dt$$
$$dy = -Axy\, dt$$

which suggests the regression model

$$CX \approx B \cdot Y0 \cdot X0 \cdot T$$
$$CY \approx A \cdot X0 \cdot Y0 \cdot T$$

so that the regression model for $CERY$ is independent of the force sizes, and hence $QYY \equiv QYX = 0$. Lanchester offers no actual combat data in support of either of his laws.

e. Lanchester-like Laws. Many of the attrition models that have been proposed since Osipov's and Lanchester's time suggest various values of the regression coefficients. One example is the logarithmic or Peterson Law, which was originally applied to the analysis of relatively small tank battles (see Peterson-1967). The differential equations for the Peterson or logarithmic law are

$$dx = -Bx\, dt$$
$$dy = -Ay\, dt$$

which suggests the regression model

$$CX \approx B \cdot X0 \cdot T$$
$$CY \approx A \cdot Y0 \cdot T$$

and the derived regression model $CERY \approx K(FRY)^{-1}$. Weiss later suggested that the Peterson or logarithmic law may be applicable to some Civil War battles (see Weiss-1966). Table 4-1 shows the regression coefficient values corresponding to the main differential equation formulations that have been proposed. However, in interpreting these laws it is crucial to bear in mind the Constant Fallacy. This is especially true when considering their regression coefficient implications.

Table 4-1. Regression Coefficients Suggested by Various Differential Equation Formulations

Law	QYY	QYX
Square	+1	−1
Osipov	+1 / 2	−1/2
Linear	0	0
Logarithmic	−1	+1

f. The Logarithmic Law. Many military organizations, specifically including the US Army (see Field Manual (FM) 101-10-1, 1987), have adopted casualty estimation methods based on an assumed constant attrition rate. Thus, they express casualty fraction rates in units of the number of casualties per 1,000 personnel-days (that is, the number of casualties per kilo-personnel-days, which we will abbreviate as /kpd). This practice suggests that the casualty numbers are jointly proportional to the size of the force and to the elapsed time, *i.e.,* that

$$CX = \alpha \cdot X0 \cdot T$$
$$CY = \beta \cdot Y0 \cdot T,$$

where α and β are constants of proportionality. Taken together, these equations imply that the casualty exchange ratio favoring side Y, $CERY \equiv CX / CY = e^{QY0}(FRY)^{-1}$, where FRY is the initial force ratio favoring side Y and $QY0$ is some constant. This regression model relating $CERY$ and FRY is the same as that suggested by Peterson's logarithmic law.

g. Early CORG Study. Helmbold-1961 uses data on 92 battles that occurred between 1741 AD and 1945 AD. In essence, he regresses the logarithm of the quantity

$$B / A = \frac{X0^2 - X^2}{Y0^2 - Y^2} \equiv \frac{(X0 - X)(X0 + X)}{(Y0 - Y)(Y0 + Y)} \approx \frac{CX}{CY} \frac{X0}{Y0}$$

on the logarithm of the quantity $X0 / Y0$. Let the resulting regression coefficients be α and β, where

$$\ln\left(\frac{CX}{CY}\frac{X0}{Y0}\right) \approx \alpha + \beta \ln\left(\frac{X0}{Y0}\right).$$

Written another way, this regression reads

$$\ln(CERY) \approx \alpha + (\beta - 1)\ln(X0 / Y0)$$

so that $QY0 = \alpha$, $QYX = (\beta - 1)$, and $QYY \equiv -QYX = (1 - \beta)$. Helmbold's results are that

$$\alpha = 0.230 \pm 0.128$$
$$\beta = 1.266 \pm 0.244$$

which corresponds to

$$QY0 = 0.230 \pm 0.128$$
$$QYX = 0.266 \pm 0.244$$
$$QYY = -0.266 \pm 0.244$$

Here, as elsewhere in this paper, the error bounds are one standard error above and below the estimated value.

h. Willard's Paper. Willard-1962 uses about 1,090 battles taken from Bodart's massive *Kriegs-Lexicon* (see Bodart-1908). Willard is forced to use the winner and loser sides, rather than the attacker and defender, since Bodart records only the winner and loser in his *Kriegs-Lexicon*. Thus, Willard's regression equations essentially read

$$\ln(CERW) = -\ln E + \gamma \ln(W0 / L0),$$

where $-\ln E$ and γ are the regression intercept and slope, respectively. Willard finds γ values ranging from -0.35 to -0.55 for Bodart Category I battles (not sieges), and from -0.27 to -0.87 for Bodart Category II battles (sieges). (See his Table 9, p 19. In each case, the ranges of values are for various subsets corresponding to different levels of total force size, timeframe, or force ratio.) These correspond to the values $QWW = -0.35$ to -0.55 for Category I battles, and to $QWW = -0.27$ to -0.87 for Category II battles. Of course, $QWL \equiv -QWW$ in either case, because that was assumed or imposed on the data from the start.

Adopting a regression model of the form

$$CERW = -\ln E + \delta \ln(W0) - \beta \ln(L0),$$

Willard finds for all 939 Category I battles (not sieges) the values $\beta = -0.21$ and $\delta = -0.52$. These correspond to the values $QWW = -0.52$ and $QWL = 0.21$ (see his Table 10, p 2). For all 149 Category II battles (sieges), he finds $\beta = -0.44$ and $\delta = -0.55$. These correspond to the values $QWW = -0.55$ and $QWL = 0.44$.

In no case does Willard report the intercept (that is, the $-\ln E$) values. Also, in his conclusions (paragraph 7, p 28), Willard observes that "...it would have been desirable to know who was the attacker in these battles...".

i. Later CORG Study. Helmbold-1964 uses data on 83 battles that occurred between 280 BC and 1944 AD. He adopts essentially the same regression model as in Helmbold-1961, and obtains a result equivalent to

$$\alpha = -0.020 \pm 0.230$$
$$\beta = 1.470 \pm 0.258$$

which corresponds to

$$QY0 = -0.020 \pm 0.230$$
$$QYX = 0.470 \pm 0.258$$
$$QYY = 0.470 \pm 0.258$$

j. Dunn's Study. Dunn-1971 uses the combined 175-battle data from Helmbold-1961 and Helmbold-1964. He adopts a regression model essentially of the form

$$\ln(CERY) = QY0 + QYY \cdot \ln(Y0) + QYX \cdot \ln(X0),$$

subject to the imposed restriction that $QYX \equiv -QYY$, and finds (see his equation 16 on p 35 and the bottom line on p 35) the values $QY0 = 0.11$ and $QYY = -0.24$.

k. Battle of Britain Study. Helmbold-1971 uses data on the air action in 18 days of the World War II Battle of Britain. Adopting essentially the same regression formulation as in Helmbold-1961 and Helmbold-1964, he obtains

$$\alpha = 0.242$$
$$\beta = 1.544 \pm 0.282$$

which corresponds to

$$QY0 = 0.242$$
$$QYX = 0.544 \pm 0.282$$
$$QYY = -0.544 \pm 0.282$$

l. Fain's Study. Fain-1977 uses 60 World War II battles drawn from an earlier version of the CDB91DAT data base. She adopts a regression model essentially of the form

$$\ln(CERY) = QY0 + QYY \cdot \ln(Y0) + QYX \cdot \ln(X0),$$

subject to the imposed restriction that $QYX \equiv -QYY$, and finds $QYY = -0.413$ or -0.594, depending on whether the sides are identified as attacker and defender or by larger and smaller (see her p 44, Notes 11 and 12).

m. Air Combat Studies. Kelley-1977 uses 138 air combat engagements from the Vietnamese War in South-East Asia (SEA) and 315 from the Korean War (NWA). He adopts a regression model essentially of the form

$$\ln(CERY) = QY0 + QYY \cdot \ln(Y0) + QYX \cdot \ln(X0),$$

where the Y side is always the US force, irrespective of important tactical features such as which side was attacking, which side was larger, or which side won. Kelley-1977 gives various figures for the coefficients, but their values tend to lie around the values $QY0 \approx 0.70$ and $QYY \approx -0.35 \approx -QYX$ for SEA and around the values $QY0 \approx 2.0$ and $QYY \approx 0.43 \approx -QYX$ for NWA.

n. Kirkpatrick's Study. Kirkpatrick-1985 uses 16 battles from Livermore's book (see Livermore-1900). He adopts a regression model of the form

$$\ln(CERY) = QY0 + QYY \cdot \ln(Y0) + QYX \cdot \ln(X0),$$

where the regression coefficients are subject to the imposed restriction $QYY \equiv -QYX$ and the sides are always Confederate (Y) and Union (X) regardless of which side attacked, won, or was stronger. He finds $QY0 = 0.01$ and $QYY \equiv -QYX = -0.99$.

o. Rowland's Study. Rowland-1987 uses data from some historical battles and from field trials. He adopts a regression model essentially of the form

$$\ln(CX) = KX0 + KXX \cdot \ln(X0) + KXY \cdot \ln(Y0),$$

where side X is the attacker and the coefficients are constrained so that $KXY \equiv 1 - KXX$. He finds $KX0 = -1.436$ and $KXY = 0.337$.

p. Dupuy's Propositions. Dupuy-1990, pp 100-103 states (among other things) the following propositions concerning the relationship of casualties to force strengths.

1. "Small force casualty rates are higher than those of large forces." This is the same as the observations of Dodge-1900, Berndt-1897, and Bodart-1908, cited above, and may be approximated by the same equations.

2. "There is no direct relationship between force ratios and attrition rates." The meaning of this not entirely clear, but it appears to say that the attrition rates RX and RY are independent of the initial force ratio, X0/Y0. It is difficult to reconcile this proposition with the immediately preceding proposition. For if the foregoing proposition is correct, it would follow that

$$\frac{RX}{RY} = FERY = FRY \times CERY = (FRY) \times K(FRY)^{-(1+PY)} X0^{PX-PY} = K(FRY)^{-PY}(X0)^{PX-PY},$$

and (except for the very special case, $PY = 0$, which is incompatible with the immediately preceding proposition), implies that attrition rates *do* depend somewhat on force ratios.

3. "In most modern battles, the numerical losses of the attacker and defender have been similar." Apart from the ambiguity of what constitutes a "modern battle," this is mathematically expressed as $CX = K \times CY$, where K is a constant of proportionality, whose value is close to unity. Because of the mathematical relationship between casualty numbers and casualty fractions, this proposition is mathematically equivalent to $FERY \equiv CERY \times FRY = K(FRY)$, where K is close to unity. This relationship is mathematically incompatible with each of the preceding propositions.

4. "In the average modern battle, attacker casualty rates are somewhat lower than defender casualty rates." Apart from the ambiguity of the phrase "average modern battle", this may be expressed mathematically as $FX = K \times FY$, where K is a constant of proportionality whose value is "somewhat" less than unity. Because of the mathematical relationship between casualty numbers and casualty fractions, this proposition is mathematically equivalent to $FERY = FX/FY = K$, where K is a constant of proportionality whose value is "somewhat" less than unity. But this implies that $CERY = FERY \div FRY = K(FRY)^{-1}$, a relationship that is mathematically incompatible with each of the immediately preceding propositions.

q. Hartley's Study. Hartley-1991 uses a set of about 860 battles obtained by combining the battles listed in Helmbold-1961 and Helmbold-1964 with those in an earlier version of the CDB91DAT data base. He adopts a regression model essentially of the form

$$\ln(CX) = KX0 + KXX \cdot \ln(X0) + KXY \cdot \ln(Y0)$$
$$\ln(CY) = KY0 + KYX \cdot \ln(X0) + KYY \cdot \ln(Y0)$$

He finds $KX0 = -4.3$, $KY0 = -3.5$, $KXX \approx KYY = 0.73 \pm 0.06$, and $KXY \approx KYX = 0.42 \pm 0.06$. Note that the values of Hartley's regression coefficients for $\ln(CX)$ are similar Rowland's (see the preceding paragraph). It appears from Hartley's values for the K's that $QY0 \approx -0.8$ and $QYY \approx -0.31 \approx -QYX$. Note, however, that in this case the relation between QYY and QYX is not imposed *a priori,* but instead is dictated by the data.

r. Summary of Past Work. Table 4-2 summarizes the prior results regarding the Q's. Here, N/U means "not used" and N/R means "used, but not reported."

As can be seen, the intercept or $QY0$ values vary quite a bit. The values for the slopes QYY and QYX obtained from actual data tend to be equal in magnitude but opposite in sign, although the significance of this is obscure because in so many cases that relationship is a constraint imposed on the data from the start. Moreover, when the sample size is sufficiently large, the value of QYX seems to rather consistently fall somewhere between +0.3 and +0.5.

Table 4-2. Summary of Prior Results for the Regression of ln*CERY* on ln*X0* and ln*Y0*

Law or Investigator	*QY0*	*QYY*	*QYX*
Square	N/U	+1	−1
Osipov	N/U	+1/2	−1/2
Linear	N/U	0	0
Logarithmic	N/U	−1	+1
Helmbold, 1961	+0.23±0.13	−0.27±0.24	+0.27±0.24
Willard, 1962	N/R	−0.55	+0.44
Helmbold, 1964	−0.02±0.23	−0.47±0.26	+0.47±0.26
Helmbold, 1971	+0.24	−0.54±0.28	+0.54±0.28
Dunn, 1971	+0.11	−0.24	+0.24
Fain, 1977	N/R	−0.41	+0.41
Kelley, 1977 (SEA)	+0.70	−0.35	+0.35
Kelley, 1977 (Korea)	+2.00	+0.43	−0.43
Kirkpatrick, 1985	+0.01	−0.99	+0.99
Hartley, 1991	−0.80	−0.31±0.06	+0.31±0.06

4-1. CRITIQUE OF PRIOR WORK. Here we offer some comments and observations on the past work.

   a. **The Constant Fallacy.** The regression results described above cannot logically be used to infer the form of differential equation governing attrition. All attempts to do so commit the pernicious Constant Fallacy, the nature of which has been amply explained in Helmbold-1994. For example, many different differential equation formulations suggest the same regression model. Hence, it is fallacious to infer *which* differential equation formulation actually corresponds to a given regression model. The fact is that such inferences can be sustained only by tacitly or openly invoking strong additional assumptions that go beyond (indeed, *well* beyond) the bounds of the regression results *per se*. The Constant Fallacy is to invoke them tacitly, but the necessary assumptions are so strong that few would dare invoke them openly. The only logically tenable method of testing the validity of various attrition differential equations is exemplified by Engel-1954. It does not depend on regression analysis for its validity. Unfortunately, lack of appropriate data has so far stymied the application of this method.

Because they commit the Constant Fallacy, remarks such as Willard's (see Willard-1962, p 20) that "Military theorists should be discouraged to find $\gamma < 0$, for in this range the results seem to imply that if the Lanchester formulation is valid the casualty-producing power of troops increases as they suffer casualties" are baseless. Fain-1977 quotes this statement with apparent approval and thus commits the Constant Fallacy. Kirkpatrick-1985 states that "The result of this investigation...provides evidence that the relationship [between losses and strengths] is much closer to Lanchester's Linear Law than to the Square Law, which is the better known and more widely quoted" and thereby commits the Constant Fallacy. None of the quoted statements are logically defensible. They all commit the Constant Fallacy of inferring a differential equation formulation from a regression model. Despite their popularity, such inferences nonetheless are a pernicious fallacy.

Helmbold-1961 and Helmbold-1964 note the empirical fact that, *if* Lanchester's Square Law is assumed to be valid for each individual battle, *then* the regression results show that the attrition coefficients must vary with the initial force ratio. This is a perfectly valid alternative interpretation of the regression results summarized above, but one which does not commit the Constant Fallacy.

Helmbold-1965 suggests yet another interpretation, namely, that forces in battle are subject to the well-known economic principle of diminishing returns to scale. According to this principle, smaller forces are more efficient than large ones. This is plausible, even when carried to the extreme of a "force" consisting of a single individual—snipers, serial killers, assassins, gangster "hit men," terrorists, and so forth notoriously have high casualty exchange ratios in their favor, even though greatly outnumbered by the society as a whole. Expanding on the original treatment in Helmbold-1965, the principle of diminishing returns to scale can lead to the following differential equations relating casualty rates to strengths, durations, and dates:

$$dx = -By(x/y)^h f(t)\,dt = -By^w x^{1-w} f(t)\,dt$$
$$dy = -Ax(y/x)^h f(t)\,dt = -Ax^w y^{1-w} f(t)\,dt$$

(4-5.1)

where $w = 1 - h$ is known as the Weiss parameter in honor of the famous military operations research analyst Herbert K. Weiss (see Taylor-1983). This is equivalent to the statement that the Lanchester's Square Law

$$dx = -\beta y\,dt$$
$$dy = -\alpha x\,dt$$

holds, except that its attrition coefficients, α and β, depend on the instantaneous force ratio according to the following equations

$$\beta = B\left(\frac{y}{x}\right)^{w-1}$$

$$\alpha = A\left(\frac{x}{y}\right)^{w-1}$$

where the coefficients A and B do not depend on x, y, or t. Note that, for small casualty levels, the instantaneous force ratio can be closely approximated by the initial force ratio. An explicit solution for these coupled nonlinear differential equations (4-5.1) is available (see Appendix D), but is not central to the point we wish to make here. Instead, we proceed as follows. From the Taylor series expansion based on (4-5.1), the casualties to side X will be approximately

$$CX = x_0 - x \approx x_0 - \left(x_0 + \frac{dx}{dt}\bigg|_0 T \right) = By_o^w x_0^{1-w} f(0)T, \qquad (4\text{-}5.2)$$

where T is the time at the end of the battle and zero subscripts indicate initial strengths. Similarly,

$$CY \approx Ax_0^w y_0^{1-w} f(0)T . \qquad (4\text{-}5.3)$$

It is tempting to infer from (4-5.2) and (4-5.3) that the regression model corresponding to (4-5.1) *must* correspond to the following coefficient values:

$$KX0 + KXD \cdot D = \ln D + \ln f(0)$$
$$KXX = 1 - w$$
$$KXY = w$$
$$KXT = 1$$
$$KY0 + KYD \cdot D = \ln A + \ln f(0) \qquad (4\text{-}5.4)$$
$$KYX = w$$
$$KYY = 1 - w$$
$$KYT = 1$$

However, this inference cannot be logically sustained—*unless* the attrition coefficients A and B, and the initial battle time factor $f(0)$, depend *only* on the date and not on any other factor that explicitly or implicitly affects casualties. That a specific differential equation relationship cannot be logically inferred from equations (4-5.4) is also explained in Helmbold-1994.

We note in passing that the Weiss parameter controls the form of the state equation as follows:

Table 4-3. Weiss Parameter and
Law

w	Law	QYY	QYX
1	Square	1	−1
3/4	Osipov	1/2	−1/2
1/2	Linear	0	0
0	Logarithmic	−1	0

b. On the Use of Battle Durations and Dates. Note that none of the prior regression analyses make use of two variables that might be considered essential—the battle duration, T, and the battle date, D. Casualties might be expected to increase as battle duration increases. And it is well known that casualties have generally declined with the passage of time. Hence, both the battle duration and battle date should be allowed to play a part in any regression analysis.

c. Use of Ordinary Least Squares Regressions. Also, we note that all of the preceding regressions used the ordinary least squares (OLS) method of regression. This is known to have attractive theoretical properties when all of the data satisfy certain important prerequisites. It is also known to have major defects when those prerequisites are not satisfied. Except for discarding a few data points that intuitively and subjectively appeared to be "outliers," little could be done about this until the relatively recent development of robust regression techniques. An excellent case can be made for the proposition that the battle data currently available for analysis fail to satisfy the prerequisites that make OLS an attractive method. Accordingly, the available data are excellent candidates for some form of robust regression.

d. Other Limitations of Prior Work. Prior work has generally failed to give the Constant Fallacy appropriate recognition, ignored battle durations and battle dates in the regression model, and depended heavily on OLS regression. In addition, prior work has commonly used a single data base (thus precluding any examination of the sensitivity of the results to different sets of data). Furthermore, prior work has commonly used only a single dependent variable (such as the casualty exchange ratio) although other dependent variables, to include the casualty numbers as well as the casualty exchange ratio, are of interest. One of the contributions of this paper is to take the first steps in redressing these shortcomings.

e. On the Choice of Regression Model. We also need to mention that (*many*) regression models relating casualties to force sizes, durations, and dates are certainly possible. In order to proceed, we found it necessary to adopt a particular model formulation. The one we finally adopted was selected as a result of judicious consideration, and after some preliminary trials using only a few alternative model formulations. These trials suggested that a model of the form we finally adopted represented the data as well or better than the alternatives that were

considered. Our choice was also guided in part by Ockham's Razor in the form of Richardson's Rapier (see Richardson-1960, p xliv): "Formulae are not to be complicated without good evidence." However, we must confess that, rather than resolving all aspects of the choice of model formulation, we have as yet barely scratched the surface.

4-6. APPROACH. This paragraph lays out our approach to the analysis of the general relations among force sizes, battle durations, and casualties. This approach seeks to advance the state of the art over prior efforts by (i) giving the Constant Fallacy appropriate recognition, (ii) using a regression model that includes the battle duration and battle date as potentially important factors, (iii), employing robust regression to minimize the distorting effects of a few gross errors in the data, (iv) systematically using more than one data base at a time in order to determine the sensitivity of the results to different sets of data, and (v) using several dependent variables, to include the casualty numbers as well as the casualty exchange ratio.

a. Regression Models. For this analysis, we adopt the regression model of paragraph 4-2d for the relation of casualty numbers to strengths, battle durations, and battle dates, and the regression model of paragraph 4-2e for the relation of casualty exchange ratios to strengths, battle durations, and battle dates.

b. Discussion of Postulated Regression Model Form

(1) Our review of prior work in paragraph 4-4 suggests that the casualty numbers in a given battle can be expected to depend on the strengths of the forces involved in the battle, and, in fact, that the logarithms of the casualty numbers are approximately linearly related to the logarithms of the strengths of the opposing sides.

(2) Observe that the chosen regression models (in effect) postulate that the principle systematic factors determining casualty numbers and casualty exchange ratios are the strengths on both sides, the battle duration, and the battle date. Note that they do not assume *a priori* that the casualties depend upon the strengths in a symmetrical fashion—for example, they do not assume *a priori* any relation between KXX, KXY, KYX, and KYY. Instead, they allow any such symmetries to be found from the empirical analysis of the data. They implicitly assume that the logarithms of the casualty numbers and of the casualty exchange ratio are approximately linearly dependent on the logarithms of the strengths of the opposing sides.

(3) Observe that the regression models adopted here treat the logarithms of the casualty numbers and casualty exchange ratio as being approximately linearly related to the battle duration and the battle date. The figures in Chapter 3 suggest that this linearity is approximately true for the battle date. The regression model implicitly assumes that this is also approximately true for the battle durations. Some preliminary analyses indicated that the form adopted here would give a somewhat better fit to the data than one that assumed the logarithms of the casualty numbers and casualty exchange ratio depended on the logarithms of the battle date and duration. The battle date has been centered at 1800 AD and normalized to a 1–century date span in order that the quantity D be of the same general magnitude as the other independent variables in the regression model.

(4) Note also that the regression models implicitly treat all factors not explicitly included as random or chance factors that contribute to the "noise" about the model regression surface, but which do not have a systematic effect upon it. Examples of the factors lumped with random noise are:

- Leadership and tactics
- Training and morale
- Terrain and fortifications
- Weather
- Logistics and transportation
- Intelligence
- Air and artillery support
- Technology

Analyses of the potential systematic influence of these other factors may be taken up in future work, but are not treated in this paper.

(5) Although the regression models we have adopted appear to be quite reasonable choices considering our present state of knowledge, they are, of course, somewhat arbitrary. Accordingly, in the long run, they should be considered as tentative approximations, subject to revision and modification—when and if such revisions are adequately supported by additional empirical data and careful analysis.

c. Some Variations Within the Regression Model Formulation. There are several variants possible even within the framework of the regression model adopted for the purposes of analysis. Some of these are mentioned below.

(1) Definition of Casualty. The first variant has to do with how casualties are defined. Shall we interpret "casualties" to mean the KIA, WIA, bloody losses (KIA + WIA), total battle casualties (TBC = KIA + WIA + CMIA), or some other measure of personnel casualties? For our work, we chose to use the nominal TBC casualties, since those are the only casualties reported by all of the data bases. (The PARCOMBO, BWSHALL, and BODASHIP data bases also report KIA, WIA, and CMIA for selected battles. Unfortunately, the data base we consider the most reliable (CDB91DAT) gives *only* TBC values, and so we adopt that as our standard.) As shown in Figure 4-1, the winner's TBC is usually pretty close to his bloody casualties, while the loser's TBC is often considerably larger than his bloody casualties—the difference being due to many loser casualties falling into the captured or missing in action category. The treatment of the bloody casualties variant is left for future work.

Table 4-4 shows a comparison between regressions done using TBC and bloody casualties. They were all done using the BWSHALL data base with OLS regression and with sieges filtered out. The column headings identify the dependent variable used, where TBCWIN = the winner's TBC, BloodWIN = the winner's bloody casualties, TBCLOS = the loser's TBC, and BloodLOS = the loser's bloody casualties. Values are rounded to the number of significant figures shown.

Table 4-4. Comparison of Regression Results Using TBC and Bloody Casualties

Value	ln(TBCWIN)	ln(BloodWIN)	ln(TBCLOS)	ln(BloodLOS)
Intercept	−2.97±0.34	−3.60±0.48	1.16±0.27	−1.22±0.33
ln(InitStrWIN)	0.68±0.05	0.72±0.07	0.18±0.04	0.15±0.05
ln(InitStrLOS)	0.33±0.05	0.37±0.07	0.52±0.04	0.73±0.05
T	0.013±0.005	0.03±0.02	0.009±0.004	0.009±0.005
D	−0.18±0.05	−0.16±0.06	−0.30±0.04	−0.37±0.05
DOF	972	379	972	632
RMSE	0.92	0.83	0.73	0.73
R^2	0.48	0.59	0.44	0.56

The column entries give the value of the regression intercept followed by the coefficients of the logarithms of the winner's and loser's strengths, battle duration (in days), and the centered and scaled date. These are followed by the number of degrees of freedom (DOF), the root mean square error of estimate (RMSE), and the square of the regression correlation coefficient (R^2). The values indicated as ± indicate plus or minus one standard error of estimate. Values are rounded to the number of significant figures shown.

As can be seen, the dependence on battle duration and battle date is practically the same regardless of whether TBC or bloody casualties are used. For the winner's casualties, only the intercept is affected. For the loser's casualties, both the intercept and the coefficient of the loser's initial strength are affected. The DOF for bloody casualties are less than those for TBC because this data base does not report bloody casualty numbers for some battles.

In our current state of knowledge it is not possible to say definitely whether TBC or bloody casualties is the "right" choice. In this paper, we use TBC, as did all of the prior work summarized in paragraph 4-4. However, interpretations and use of the results should take cognizance of the fact that we standardized on TBC in our analyses.

(2) Definition of Strengths. Another variant has to do with how personnel strength is to be defined. Shall we interpret strength as the initial strength (a practice suggested by the fact that the differential equations (4-1) contain no reinforcement terms), the total number of personnel *committed* to action (including reinforcements actually committed to action), the total number of personnel *available* for commitment to action (including those held in reserve without being committed to action), or some other measure of personnel strength? In a practical sense this distinction is meaningful only for the CDB91DAT data base—none of the others say exactly what their strength figures represent. For our work, we interpret strength as the initial strength and filter out any of the CDB91DAT battles where it was known that significant reinforcements

CAA-RP-95-1

were used. The following Table 4-5 illustrates the differences among regression coefficients using different definitions of initial strengths. All of these regressions were done with the CDB91DAT data base using OLS regression and taking either ln(CASA) or ln(CASD) as the dependent variable. Values are rounded to the number of significant figures shown.

Table 4-5. Comparison of Regressions Using Initial or Total Force Size

Value	ln(CASA) Type A	ln(CASA) Type B	ln(CASD) Type A	ln(CASD) Type B
Intercept	−1.93±0.49	−2.51±0.44	−1.22±0.53	−1.67±0.46
ln(INTSTA)	+0.73±0.10	+0.75±0.09	+0.35±0.11	+0.35±0.09
ln(INTSTD)	+0.21±0.09	+0.25±0.08	+0.53±0.10	+0.59±0.09
T	+0.06±0.05	+0.07±0.03	+0.11±0.06	+0.10±0.03
D	−0.39±0.07	−0.36±0.06	−0.47±0.07	−0.44±0.07
DOF	289	347	289	347
RMSE	1.07	1.05	1.15	1.11
R^2	0.61	0.64	0.60	0.63

Here the column headings indicate the dependent variable and the type of regression used. Type A uses the initial strengths and filters out all battles whose reinforcements are not known. Type B uses the total strengths and does not filter out any battles. As can be seen, the use of different strengths does not appreciable change the coefficients of the independent variables. Although there is a change in the value of the intercept, it is not easy to tell whether this is due solely to the different definitions of strengths, solely to the use of a different set of battles, or to a complex interaction of these two effects. A more detailed treatment of the effects of committed or available reinforcements is left to future work.

(3) **Definition of Battle Durations.** Another variant has to do with how battle durations are defined. Shall we interpret the durations as the active durations, elapsed durations, overall durations, or some other measure of the duration? Based on the results of Chapter 3, in most battles there is little difference between the elapsed and active time figures. In our work, we use the active times whenever they are available, for those are the times during which most of the casualties occur. However, in order to be as complete as possible we shall also use elapsed and overall durations when these are available.

(4) **Definition of Battle Dates.** Another variant has to do with the choice of scale used to define dates. Shall we use the conventional Anno Domini dates and the year as the unit of elapsed calendar time, or should we use some other scale? We elected to use dates converted to a scale in which the "zero year" is 1800 AD and the unit of elapsed calendar time is the century. This choice was made based on a desire to scale the dates so that they had about the same range of variation as the other independent variables in the regression model.

(5) Definition of Sides. Shall we use attacker and defender, winner and loser, or some other definition of side? In our work, we used the attacker and defender sides whenever available. However, in order to be as complete as possible, as well as to maintain comparability with the BWSHALL and BODASHIP data bases which report only winner and loser sides, we also make use of winner and loser sides whenever these are available.

(6) Choice of Regression Method. Shall we use ordinary least squares fits to the data, employing conventional ordinary least squares (OLS) regression computations? This method is commonly taught and widely known. It has excellent theoretical properties whenever the requisite assumptions are satisfied. However, it is well known to be strongly affected by erroneous data. A wide variety of robust regression procedures designed to be substantially less sensitive to erroneous data, yet retain most of the advantages of the OLS method, are also available. In order to gain some protection against gross errors and mistakes in the data, we adopted a robust regression method. The specific robust regression method adopted is known as Andrew's Sine, with a tuning constant of 2.1 and iterated until there is less than 1 percent change in the regression coefficients and mean square error. (The tuning constant value of 2.1 is suggested in the literature as being close to the best compromise between discounting valid data points and giving too much weight to invalid ones.)

Table 4-6 shows a comparison of the results obtained using OLS and two alternative robust regression methods. They were all done using the PARCOMBO data base with ln(CERD) as the dependent variable, attacker and defender sides, active durations, and filtered to omit battles prior to 1500 AD. The notations in parentheses are: OLS = ordinary least squares regression used, S11 = Andrew's sine robust regression with tuning constant 2.1 and iterated until there is less than a 1 percent change to any regression coefficient and to the RMSE, T11 = Tukey's biweight robust regression with tuning constant 6.0 and iterated until there is less than a 1 percent change to any regression coefficient and to the RMSE. Values are rounded to the number of significant figures shown. (The tuning constant values used are those suggested in the literature as being close to the best compromise between discounting valid data points and giving too much weight to invalid ones.)

**Table 4-6. Comparison of OLS and Robust
Regression Results**

Value	lnCERD (OLS)	lnCERD (S11)	lnCERD (T11)
Intercept	−2.82±0.71	−0.64±0.48	−1.44±0.60
ln(INTSTA)	+0.76±0.11	+0.32±0.09	+0.34±0.11
ln(INTSTD)	−0.50±0.10	−0.25±0.07	−0.19±0.09
ACTDURN	−0.01±0.01	−0.01±0.01	−0.01±0.01
DATECENT	+0.02±0.13	−0.01±0.08	+0.06±0.10
DOF	131	130	130
RMSE	0.94	0.48	0.67
R^2	0.26	0.10	0.08

As can be seen, the robust regression results are in reasonable accord with each other, but differ appreciably from the OLS results. From the DOF, it appears that the robust regression discounted only a small number of erroneous points that threw off the OLS regression results.

Table 4-7 illustrates how the estimated regression coefficients change as the tuning constant is varied. These regressions were all done using the PARCOMBO data base with lnCERD as the dependent variable, attacker and defender sides, active durations, filtered to omit all battles prior to 1500 AD, Andrew's sine robust regression, and various tuning constants (tuning constant values given by the column headings). Values are rounded to the number of significant figures shown.

**Table 4-7. Comparison of Robust Regression Results Using Various
Tuning Constants**

Value	1.0	1.5	2.1	3.0	4.0	5.0
Intercept	+0.04±0.48	−0.56±0.48	−0.64±0.48	−0.64±0.48	−0.66±0.48	−0.66±0.48
ln(INTSTA)	+0.03±0.09	+0.29±0.09	+0.32±0.09	+0.32±0.09	+0.33±0.09	+0.33±0.09
ln(INTSTD)	−0.02±0.07	−0.23±0.07	−0.25±0.07	−0.25±0.07	−0.26±0.07	−0.26±0.07
ACTDURN	−0.00±0.01	−0.01±0.01	−0.01±0.01	−0.01±0.01	−0.01±0.01	−0.01±0.01
DATECENT	+0.13±0.07	−0.02±0.08	−0.01±0.08	−0.01±0.08	−0.02±0.08	−0.02±0.08
DOF	116	128	130	130	131	131
RMSE	0.44	0.47	0.48	0.48	0.48	0.48
R^2	0.03	0.09	0.10	0.10	0.11	0.11

Table 4-7 illustrates that a tuning constant value less than 2.1 tends to discount too many valid data points. However, little is gained by increasing the tuning constant above 2.1. So these results tend to confirm the tuning constant values recommended in the literature, and support their applicability to our analyses.

(7) Choice of Data Bases. What specific data should be used? We elected to use the CDB91DAT, PARCOMBO, BWSHALL, and BODASHIP data bases. However, in most cases, we filtered these data bases to make the data actually used in the regression analyses as comparable as possible. In particular, we filtered the CDB91DAT to omit all battles with reinforcements. We also filtered the PARCOMBO data to omit a handful of battles before 1500 AD because the other data bases included only battles later than 1600 AD. In effect, this discounts the data from battles before that date. The BWSHALL data were filtered to omit sieges because the other data bases have practically no sieges, and so it would not be proper to include them in the BWSHALL data. (Some preliminary analyses indicated that sieges were sufficiently different from the other types of battles that they ought to be treated separately, rather than being lumped together.) The BODASHIP data were not filtered in any way, but, because they are all naval engagements, they may not be comparable to land battles.

(8) Summary of Approach. The foregoing considerations can be summarized as follows.

(a) The independent variables chosen for the regression model are the logarithms of the opposing strengths ($\ln X0$ and $\ln Y0$, or $\ln W0$ and $\ln L0$), the battle duration (T), and the centered and scaled battle date (D). The dependent variables are the logarithms of the total battle casualty numbers ($\ln CX$ and $\ln CY$, or $\ln CW$ and $\ln CL$), the logarithm of the casualty number exchange ratio favoring the defender or the winner ($\ln CERY$ or $\ln CERW$), and the logarithm of the fraction exchange ratio favoring the defender or the winner ($\ln FERY$ or $\ln FERW$).

(b) The regressions are to be done using various data bases, as follows.

- CDB91DAT (the CDB90 data base of battles generated during PAR-Phase 1, filtered to omit all battles with reinforcements to either side).

- PARCOMBO (a composite data base created from components developed during PAR-Phase 1, filtered to omit all battles that occurred prior to 1500 AD).

- BWSHALL (the Bodart land battle data base generated during PAR-Phase 1, filtered to omit sieges).

- BODASHIP (Bodart naval battle data base, no filtering).

(c) The regression method to be used is Andrew's sine robust regression with a tuning constant of 2.1 and iterations continued until there is less than a 1 percent change either to any regression coefficient or to the RMSE.

CAA-RP-95-1

(d) Various combinations of durations and sides are used with the various data bases, as shown in Table 4-8 below. In this table, the durations are Overall (whole day durations), Elapsed (from the time the battle started to the time it ended), or Active (sum of the active time intervals). Also, the sides are ATK/DEF (attacker and defender) or WIN/LOS (winner and loser).

As can be seen, there are 12 data base/duration/side cases to be considered. With each of these, we are to consider each of four dependent variables (these are either $\ln CX$, $\ln CY$, $\ln(CERY)$, and $\ln(FERY)$, or else the corresponding variables when the winner and loser sides are used). Thus, there are 12 x 4 = 48 cases in all that need to be analyzed in detail.

Table 4-8. Summary of Regression Approach

No.	Side type (A/D = ATK/DEF W/L = Win/LOS)	Data base	Duration type
1	A/D	CDB91DAT[1]	Overall
2	ditto	ditto	Elapsed
3	ditto	ditto	Active
4	ditto	PARCOMBO[2]	Overall
5	ditto	ditto	Active
6	W/L	CDB91DAT[1]	Overall
7	ditto	ditto	Elapsed
8	ditto	ditto	Active
9	ditto	PARCOMBO[2]	Overall
10	ditto	ditto	Active
11	ditto	BWSHALL[3]	Overall
12	ditto	BODASHIP	Overall

[1] Filtered to omit battles with reinforcements to either side.
[2] Filtered to omit battles that occurred prior to 1500 AD.
[3] Filtered to omit sieges.

4-7. REGRESSION RESULTS WHEN ATTACKER AND DEFENDER SIDES ARE USED.

a. This paragraph describes the regression results obtained when attacker and defender sides are used. A later paragraph will describe the regression results when winner and loser sides are used. Thus, for this paragraph, only the CDB91DAT and PARCOMBO data bases can be used. This paragraph displays the results graphically. A tabular presentation of the results is contained in Appendix D.

b. Figures 4-2 through 4-5 show, for various data bases and choices of battle duration, the regression coefficients when the logarithm of the attacker's casualty number, $\ln CX$, is regressed on the logarithms of the opposing strengths ($\ln X0$ and $\ln Y0$), the battle duration (T), and the centered and scaled battle date (D). These figures graphically show the estimated value of the regression coefficient (identified as Est) and its associated plus–or–minus one standard error uncertainty band (identified as Est+SE and Est–SE). It is apparent that the regression coefficients for the CDB91DAT and PARCOMBO data bases differ significantly. The major difference is in the rate at which the logarithm of the attacker's casualty number ($\ln CX$) is declining with respect to battle date. The CDB91DAT data indicate that the decline with respect to battle date is quite rapid, and this is paired with a constant term that is only slightly negative. On the other hand, the PARCOMBO data indicate that the decline with respect to battle date is much more gradual, but this is paired with a constant term that is strongly negative.

The net effect of this difference is that the CDB91DAT and PARCOMBO regressions differ substantially at earlier dates, but agree more closely for present and future dates. This can be illustrated by considering the regression coefficients when battle durations are taken to be their active durations. Then the date–dependent effective constant terms for the PARCOMBO and CDB91DAT data will be

$$CX0(DateCent) = -0.464 - 0.596 DateCent$$
$$CX0(DateCent) = -1.638 - 0.154 DateCent$$

respectively, where the values given are from the regression computations rounded to three significant figures, rather than values read off the graphs. Table 4-9 illustrates the difference between these two equations.

Table 4-9. Comparison of Effective Constant Terms for Regression of $\ln CX$ Using PARCOMBO and CDB91DAT

	C0(DateCent)	
Date	PARCOMBO	CDB91DAT
1500	−1.18	1.32
1600	−1.33	0.73
1700	−1.49	0.13
1800	−1.64	−0.46
1900	−1.79	−1.06
2000	−1.95	−1.66
2100	−2.10	−2.25

As can be seen from this table, the date–dependent effective constant term for the CDB91DAT data start very high, but drop steeply. On the other hand, the date–dependent constant term for the PARCOMBO data start lower, but decline less steeply. The net effect is that they approach

each other, and become equal somewhere around 2070 AD. The exact date of the crossover is, of course, highly uncertain and not exactly determinable from the available data. The important point is that projections of casualty numbers to the near future will be about the same, whether those projections are based on the CDB91DAT or the PARCOMBO data.

c. Figures 4-6 through 4-9 show for various data bases and choices of battle duration the regression coefficients when the logarithm of the defender's casualty number, $\ln CY$, is regressed on the logarithms of the opposing strengths ($\ln X0$ and $\ln Y0$), the battle duration (T), and the centered and scaled battle date (D). These figures graphically show the estimated value of the regression coefficient (identified as Est) and its associated plus-or-minus one standard error uncertainty band (identified as Est+SE and Est-SE). It is apparent that the regression coefficients for the CDB91DAT and PARCOMBO data bases differ significantly. One of the major differences is again in the rate at which the logarithm of the casualty number ($\ln CY$) declines with respect to battle date. The CDB91DAT data indicate a steeper decline with respect to battle date than the PARCOMBO data. The effective date-dependent constant terms for PARCOMBO and CDB91DAT are

$$CY0(DateCent) = -0.516 - 0.430 DateCent$$
$$CY0(DateCent) = -0.954 - 0.519 DateCent$$

respectively. Table 4-10 illustrates the difference between these two equations. However, in this case the crossover point is in the early 1300s. Other differences are in the behavior of the coefficients of $\ln X0$ and $\ln Y0$, with the CDB91DAT data showing these coefficients to be about the same and having little change when the battle duration is changed from Active to Elapsed to Overall. On the other hand, the PARCOMBO data show a substantial difference between the coefficients of $\ln X0$ and $\ln Y0$, especially for the Active duration case.

Table 4-10. Comparison of Effective Constant Terms for Regression of ln*CY* Using PARCOMBO and CDB91DAT

	C0(DateCent)	
Date	PARCOMBO	CDB91DAT
1500	0.77	0.61
1600	0.34	0.09
1700	−0.09	−0.43
1800	−0.52	−0.95
1900	−0.95	−1.47
2000	−1.38	−1.99
2100	−1.81	−2.51

d. Figures 4-10 through 4-13 show, for various data bases and choices of battle duration, the regression coefficients when the logarithm of the casualty exchange ratio favoring the defender, $\ln CERY$, is regressed on the logarithms of the opposing strengths ($\ln X0$ and $\ln Y0$), the battle duration (T), and the centered and scaled battle date (D). These figures graphically show the estimated value of the regression coefficient (identified as Est) and its associated plus–or–minus one standard error uncertainty band (identified as Est+SE and Est–SE). Here, there is substantial agreement between the CDB91DAT and PARCOMBO data bases for all choices of battle duration. The regression coefficients are nearly the same whichever combination of data base and battle duration is used.

Note that the regression coefficients of T and D are nearly zero, indicating that the battle duration and battle date have little direct influence on the casualty exchange ratio favoring the defender, $CERY$. Note, however, that there may be a very gradual decline in $\ln CERY$ with respect to battle date, amounting to something in the neighborhood of -0.025 per century.

Also note that the constant term is negative, generally ranging from -0.5 to -0.75. This indicates that, all other things being equal, the casualty exchange ratio favoring the defender is less than 1, that is, that the defender is at a rather sizable *dis*advantage relative to the attacker.

The error bounds on the regression coefficients of T and D are generally much smaller for the PARCOMBO than for the CDB91DAT data. The reasons for this are obscure.

The pattern of results for the coefficients of $\ln X0$ and $\ln Y0$ is intriguing. As Figure 4-11 shows, they are nearly mirror images of each other, except perhaps for a very slight positive bias. This bias is illustrated in Figure 4-14, which shows the sum of the regression coefficients of $\ln CERY$ on $\ln X0$ and $\ln Y0$, that is, $QYY + QYX$. Here the standard errors have been computed conservatively, that is, ignoring any possible correlation between the estimated coefficient values. The resulting value of the standard error tends to err on the high side. Nevertheless, it appears that a defensible position is that the values of these regression coefficients are equal in value, but opposite in sign. Based on Figure 4-11, 0.40 is a suitable nominal figure for the common value.

Based on the above results, we take the coefficients $QYY = -0.40$, $QYX = +0.40$ and take the other coefficients to be zero. Based on Figure 4-10, a reasonable value for the constant term $QY0$ is $-0.693 = \ln(1/2)$. In other words, a reasonable approximation to the relationship of $CERY$ to force ratio is

$$CERY \approx \frac{1}{2}(FRY)^{-0.40}.$$

The signs of the coefficients of $\ln X0$ and $\ln Y0$ in Figure 4-11, and the equation in the preceding paragraph, indicate that the casualty exchange ratio favoring the defender *decreases* as the force ratio favoring the defender *increases*. It might have been thought that the effect would be opposite to this, that is, that the casualty exchange ratio favoring the defender would *increase* as the force ratio favoring the defender *increases*. But the data are quite clear that this is not what happens. The explanation presumably is the phenomenon of decreasing returns to scale. In

other words, small forces tend to be more efficient than large ones. For example, a 20,000–man division opposed by a 200–man company would have great difficulty in bringing *all* of its resources to bear effectively on its opponent, while the 200–man company would have a much better opportunity to make full use of all its resources. As a result, the casualty exchange ratio favoring this 200–man company would be much higher than anticipated in view of the very small force ratio in its favor. It may be that Clausewitz had this phenomenon at least partly in mind when he referred to "friction" as a general factor reducing the efficiency of all military organizations. The principle of economy of force also seems to be intimately related to this phenomenon.

    e. When the logarithm of the fractional exchange ratio favoring the defender, $\ln FERY$, is regressed on the logarithms of the opposing strengths ($\ln X0$ and $\ln Y0$), the battle duration (T), and the centered and scaled battle date (D), the results are essentially the same as when the logarithm of the casualty exchange ratio favoring the defender, $\ln CERY$, is regressed on the same variables. The only difference is that the coefficients of $\ln X0$ and $\ln Y0$ are the complementary values. That is, if the regression of $\ln CERY$ gives the coefficient values $+0.40$ for $\ln X0$ and -0.41 for $\ln Y0$, then the regression of $\ln FERY$ will give the coefficient values $-0.60 = +0.40 - 1.00$ for $\ln X0$ and $+0.59 = +1.00 - 0.41$ for $\ln Y0$. All the other regression coefficients for $\ln FERY$, including the constant term, will be identical to those for $\ln CERY$. The reason for this is the mathematical relation between these two quantities, namely

$$\ln FERY = \ln CERY + \ln FRY = \ln CERY + \ln(Y0 \,/\, X0),$$

together with the fact that exactly the same set of battles is used for both of these regressions (cf. Appendix D).

    Accordingly, based on the results of the regression for $\ln CERY$, we can state the following findings regarding the regression of the logarithm of the fractional exchange ratio favoring the defender, $\ln FERY$, on the logarithms of the opposing strengths ($\ln X0$ and $\ln Y0$), the battle duration (T), and the centered and scaled battle date (D). First, there is substantial agreement between the CDB91DAT and PARCOMBO data bases for all choices of battle duration. The regression coefficients are nearly the same whichever combination of data base and battle duration is used. Second, the regression coefficients of T and D are nearly zero, indicating that the battle duration and battle date have little direct influence on the fractional exchange ratio favoring the defender, *FERY*. Note, however, that there may be a very gradual decline in $\ln FERY$ with respect to battle date, amounting to something in the neighborhood of -0.025 per century. Third, the constant term is negative, generally ranging from -0.5 to -0.75, so, all other things being equal, the fractional exchange ratio favoring the defender is less than one, that is, that the defender is at a rather sizable *dis*advantage relative to the attacker. Fourth, the pattern of results for the coefficients of $\ln X0$ and $\ln Y0$ is intriguing. They are nearly mirror images of each other, except for a very slight positive bias. Nevertheless, it appears that a defensible position is that the values of these regression coefficients are equal in value, but opposite in sign. A value of 0.60 is a suitable nominal figure for this common value. Thus, the suggested approximate relationship between *FERY* and force ratio is

$$FERY \approx \frac{1}{2}(FRY)^{0.60}$$

Finally, we point out that the signs of the regression coefficients for the logarithm of the fractional exchange ratio favoring the defender, ln *FERY*, show that this quantity does increase as the force ratio favoring the defender increases (although at a rate substantially less than directly proportional, in accord with the principle of diminishing returns to scale). This behavior, among other considerations (see Helmbold–1986), suggests that the fractional exchange ratio favoring the defender is a more satisfactory index of victory in battle than the casualty exchange ratio favoring the defender.

4-8. REGRESSION RESULTS WHEN WINNER AND LOSER SIDES ARE USED

a. This paragraph describes the regression results obtained when winner and loser sides are used. Thus, for this paragraph, we make use of the BWSHALL and BODASHIP data bases, as well as the CDB91DAT and PARCOMBO data bases. This paragraph displays the results graphically. A tabular presentation of the results is contained in Appendix D.

b. Figures 4-15 through 4-18 show, for various data bases and choices of battle duration, the regression coefficients when the logarithm of the winner's casualty number, ln *CW*, is regressed on the logarithms of the opposing strengths (ln *W0* and ln *L0*), the battle duration (T), and the centered and scaled battle date (D). These figures graphically show the estimated value of the regression coefficient (identified as Est) and its associated plus or minus one standard error uncertainty band (identified as Est+SE and Est–SE). It is apparent that the regression coefficients differ significantly for the various combinations of data base and choice of battle duration. The Overall BWSHALL and Overall BODASHIP cases seem to be particularly at variance with the other cases.

c. Figures 4-19 through 4-22 show for various data bases and choices of battle duration the regression coefficients when the logarithm of the loser's casualty number, ln *CL*, is regressed on the logarithms of the opposing strengths (ln *W0* and ln *L0*), the battle duration (T), and the centered and scaled battle date (D). These figures graphically show the estimated value of the regression coefficient (identified as Est) and its associated plus–or–minus one standard error uncertainty band (identified as Est+SE and Est–SE). It is apparent that the regression coefficients differ significantly for the various combinations of data base and choice of battle duration.

d. Figures 4-23 through 4-26 show for various data bases and choices of battle duration the regression coefficients when the logarithm of the casualty exchange ratio favoring the winner, ln *CERW*, is regressed on the logarithms of the opposing strengths (ln *W0* and ln *L0*), the battle duration (T), and the centered and scaled battle date (D). These figures graphically show the estimated value of the regression coefficient (identified as Est) and its associated plus or minus one standard error uncertainty band (identified as Est+SE and Est–SE). It is apparent that the regression coefficients differ significantly for the various combinations of data base and choice of battle duration. However, there is general agreement among the various data bases and duration choices regarding certain important points. For example, Figures 4-25 and 4-26 indicate

that battle duration and battle date do not have a particularly large influence on the casualty exchange ratio favoring the winner.

Here the constant term is strongly positive, generally ranging upward of +1.5. This indicates that, all other things being equal, the casualty exchange ratio favoring the winner is quite large compared to 1. Thus, all other things being equal, the winner has a very substantial advantage relative to the loser. This is, of course, just what one would expect—the winner generally inflicts proportionately greater casualty numbers than the loser.

Also, Figure 4-24 shows that the coefficients of ln $W0$ and ln $L0$ are at least roughly mirror images of each other. That is, they are roughly numerically equal but opposite in sign, except for a possible slight negative bias. Figure 4-27 shows the sum of these coefficients. A figure of about –0.1 appears to be an acceptable nominal value for the bias, at least for the CDB91DAT data. However, if we suppress the bias term, these coefficients would each have a numerical value close to 0.5. If we suppress the bias term, this suggests the following rough approximation to the relationship between $CERW$ and force ratio

$$CERW \approx 7.4(FRW)^{-0.5}$$

The signs of the coefficients of ln $W0$ and ln $L0$ shown on Figure 4-24, and the equation in the preceding paragraph, indicate that the casualty exchange ratio favoring the winner *decreases* as the force ratio favoring the winner *increases*. It might have been thought that the effect would be opposite to this, that is, that the casualty exchange ratio favoring the winner would *increase* as the force ratio favoring the winner *increases*. But the data are quite clear that this is not what happens. This result presumably is due to the phenomenon of decreasing returns to scale.

e. When the logarithm of the fractional exchange ratio favoring the winner, ln $FERW$, is regressed on the logarithms of the opposing strengths (ln $W0$ and ln $L0$), the battle duration (T), and the centered and scaled battle date (D), the results are essentially the same as when the logarithm of the casualty exchange ratio favoring the winner, ln $CERW$, is regressed on the same variables. The only difference is that the coefficients of ln $W0$ and ln $L0$ are the complementary values. That is, if the regression of ln $CERW$ gives the coefficient values +0.47 for ln $L0$ and –0.55 for ln $W0$, then the regression of ln $FERW$ will give the coefficient values –0.53 = +0.47 – 1.00 for ln $L0$ and +0.45 = +1.00 – 0.55 for ln $W0$. All the other regression coefficients for ln $FERW$, including the constant term, will be identical to those for ln $CERW$. The reason for this is the mathematical relation between these two quantities, namely

$$\ln FERW = \ln CERW + \ln FRW = \ln CERW + \ln(W0 / L0),$$

together with the fact that exactly the same set of battles is used for both of these regressions (cf. Appendix D).

Accordingly, based on the results for the regression for ln $CERW$, we can state the following findings regarding the regression of the logarithm of the fractional exchange ratio favoring the winner, ln $FERW$, on the logarithms of the opposing strengths (ln $W0$ and ln $L0$), the battle

duration (T), and the centered and scaled battle date (D). First, there is rough agreement among the various data bases for all choices of battle duration. The regression coefficients are roughly the same whichever combination of data base and battle duration is used. Second, the regression coefficients of T and D are small, indicating that the battle duration and battle date have only a weak direct influence on the fractional exchange ratio favoring the winner, *FERW*. Third, the constant term is large and positive, generally ranging upwards of +1.5, which indicates that, all other things being equal, the fractional exchange ratio favoring the winner is quite large compared to one, and the winner has a very substantial advantage relative to the loser. Fourth, the pattern of results for the coefficients of $\ln W0$ and $\ln L0$ is intriguing. They are roughly mirror images of each other, except for a small negative bias.

Finally, we point out that the signs of the regression coefficients for the logarithm of the fractional exchange ratio favoring the winner, $\ln FERW$, show that this quantity does increase as the force ratio favoring the winner increases. (Although at a rate less than directly proportional, in accord with the principle of diminishing returns to scale.) This behavior, among other considerations (see Helmbold–1986), suggests that the fractional exchange ratio favoring the winner is a more satisfactory index of victory in battle than the casualty exchange ratio favoring the winner.

4-9. SUMMARY OF FINDINGS

a. This paragraph summarizes the findings of this chapter.

b. It is not true that the defender has some inherent advantages over the attacker. In fact, the attacker has generally taken fewer TBC than the defender. On the average, the casualty exchange ratio favoring the defender is less than 1, and the fractional exchange ratio favoring the defender is also less than 1.

c. Smaller forces take and inflict proportionately more casualties than larger forces. It is conjectured that this is a result of diminishing returns to scale.

d. Neither Lanchester's square law, Osipov's law, nor Peterson's logarithmic law are good approximations to the true relation of casualty fractions to force ratios.

e. The data do not reveal the expected (that is, approximately linear) dependency of casualty numbers on the temporal duration of a battle. One might expect that the casualty numbers would be in direct proportion to the duration of a battle, but this is clearly not what the data show.

f. The casualty exchange ratio favoring the defender *decreases* as the force ratio favoring the defender *increases*, despite what one might have expected. In fact, an approximate equation relating *CERY* to *FRY* is

$$CERY \approx \frac{1}{2}(FRY)^{-0.4}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

g. However, the fractional exchange ratio favoring the defender does increase as the force ratio favoring the defender increases. An approximate equation relating *FERY* to *FRY* is

$$FERY \approx \frac{1}{2}(FRY)^{+0.6}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

h. The casualty exchange ratio favoring the winner *decreases* as the force ratio favoring the winner *increases*, despite what one might have expected. In fact, an approximate equation relating CERW to FRW is

$$CERW \approx 7.4(FRW)^{-0.5}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

i. However, the fractional exchange ratio favoring the winner does increase as the force ratio favoring the winner increases. An approximate equation relating *FERW* to *FRY* is

$$FERW \approx 7.4(FRW)^{+0.5}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

j. Other casualty relations appear to differ, depending on the data base use, the battle duration, and the battle date. The reasons for these differences are left to future investigations.

BWSHALL: Bloody Casualties vs TBC
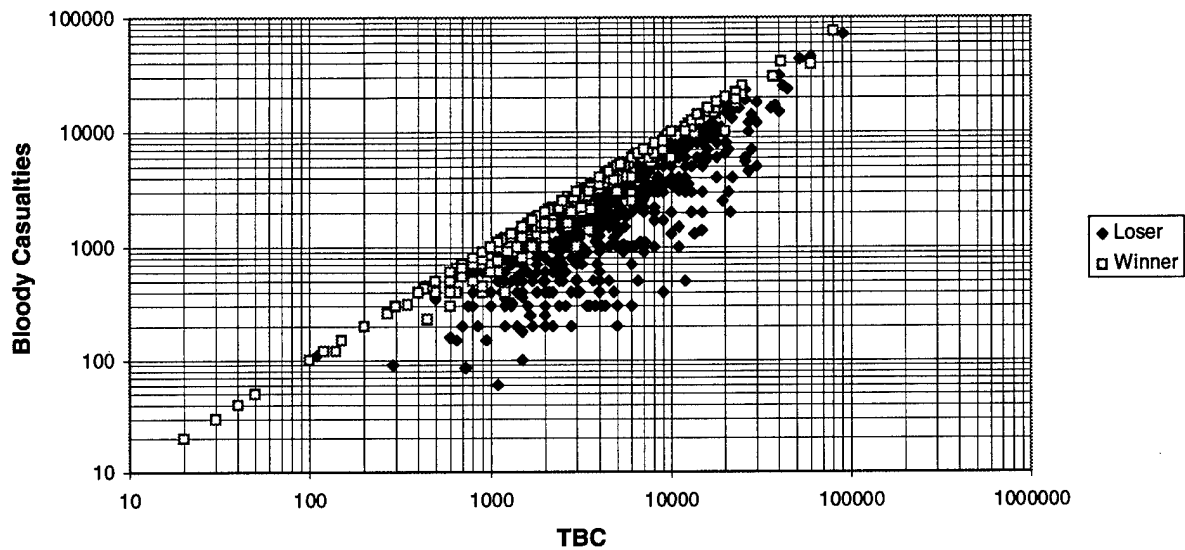
Figure 4-1. Bloody versus TBC Casualties for the Winner and Loser for the BWSHALL Data Base

A/D lnCX, Constant Term
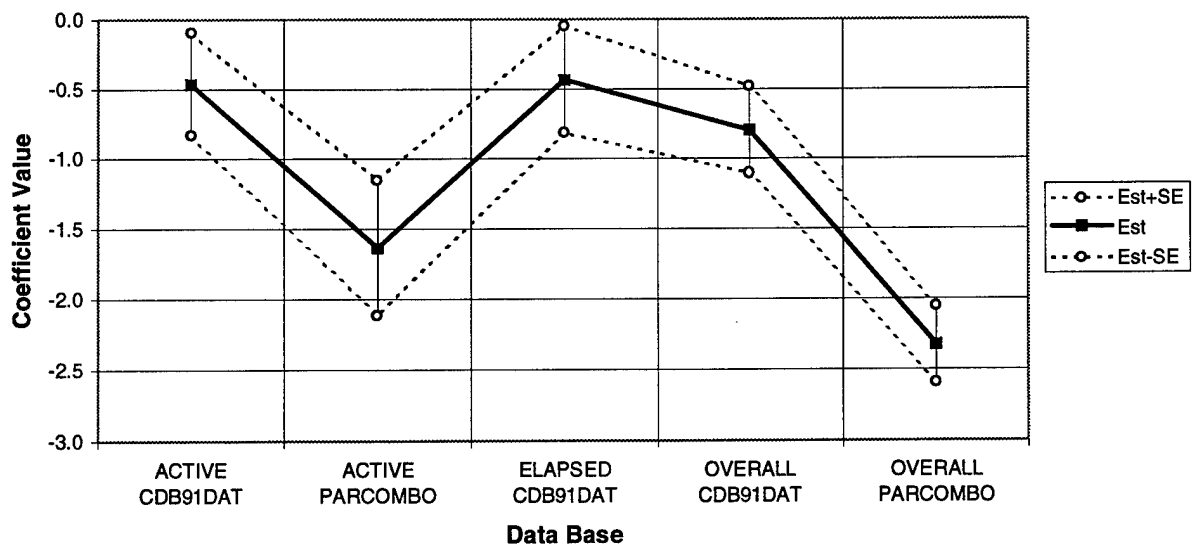
Figure 4-2. Constant Term for Regression of lnCX using Various Data Bases

A/D lnCX, lnX0 and lnY0 Coefficients
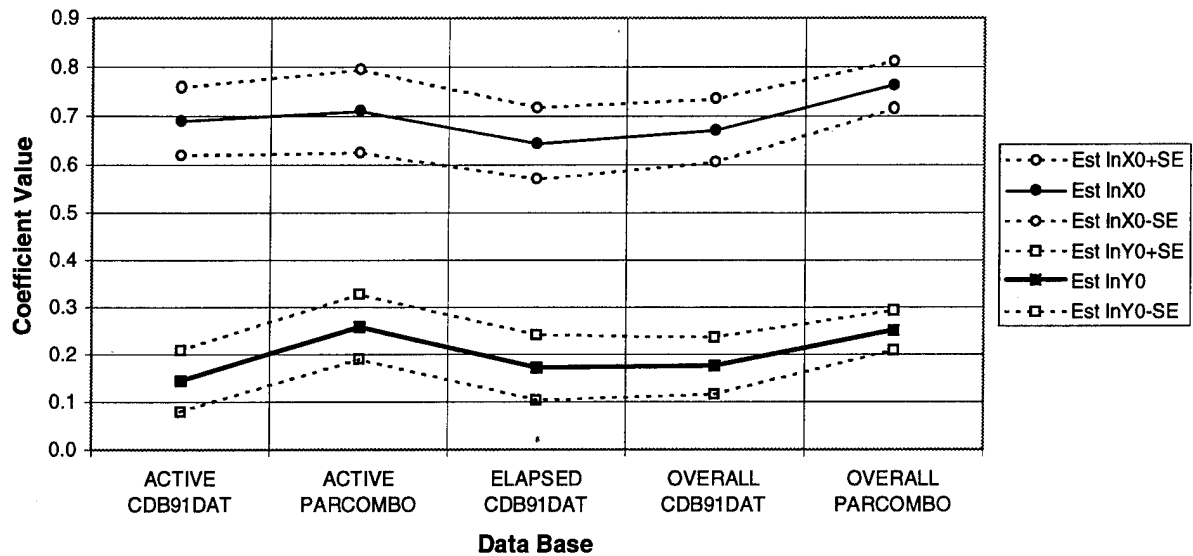
Figure 4-3. Coefficients of lnX0 and lnY0 for Regression of lnCX using Various Data Bases

A/D lnCX, Coefficient of T
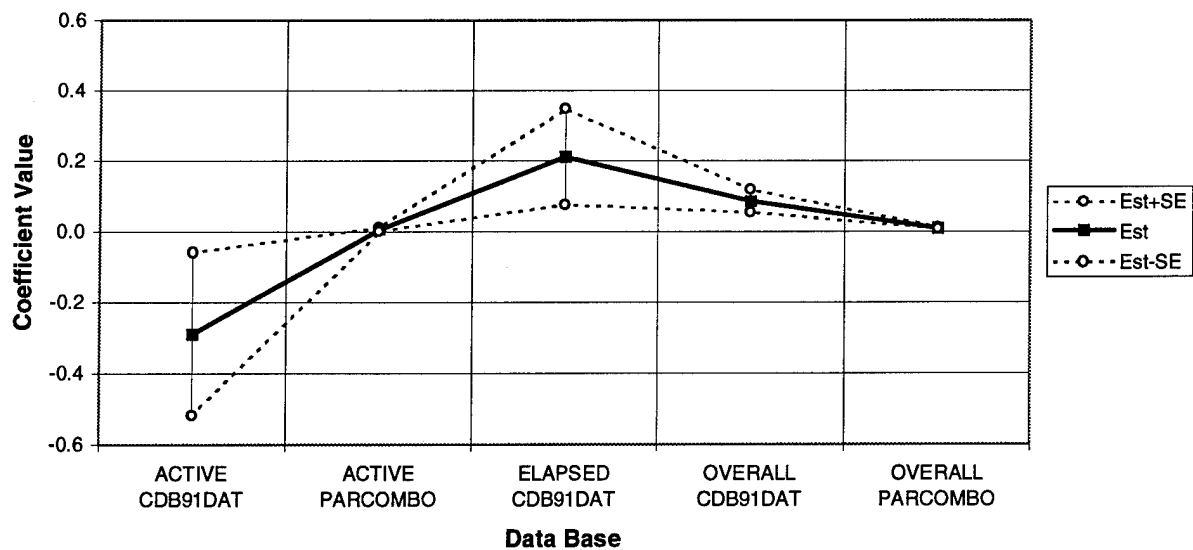
Figure 4-4. Coefficient of T for Regression of lnCX using Various Data Bases

A/D lnCX, Coefficient of D
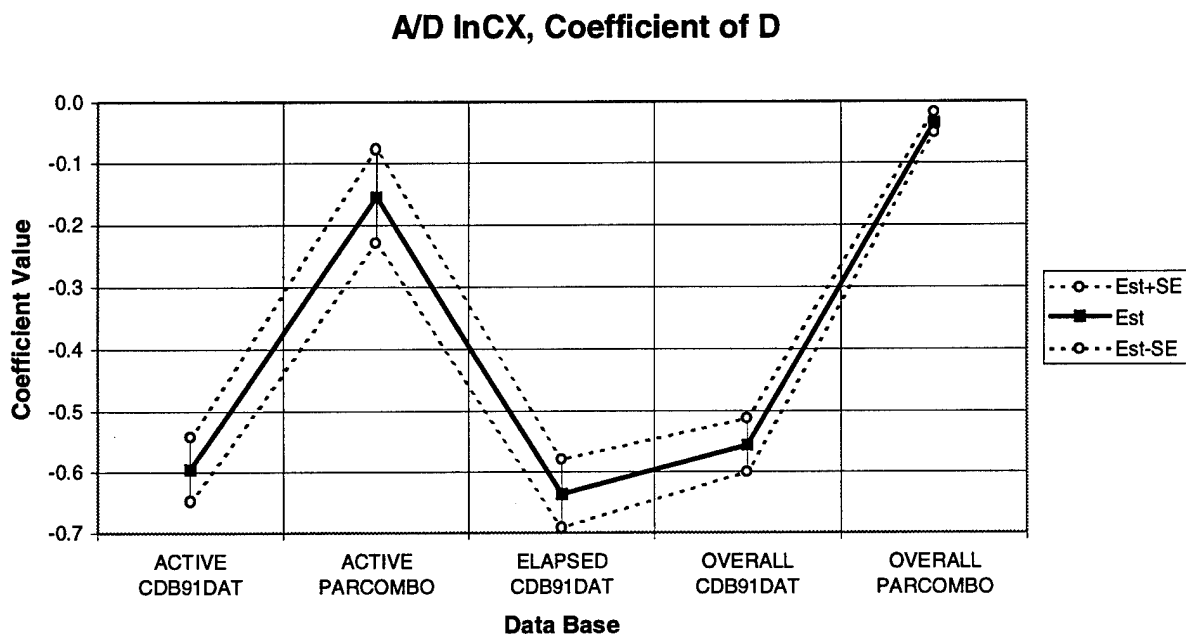
Figure 4-5. Coefficient of D for Regression of lnCX using Various Data Bases

A/D lnCY, Constant Term
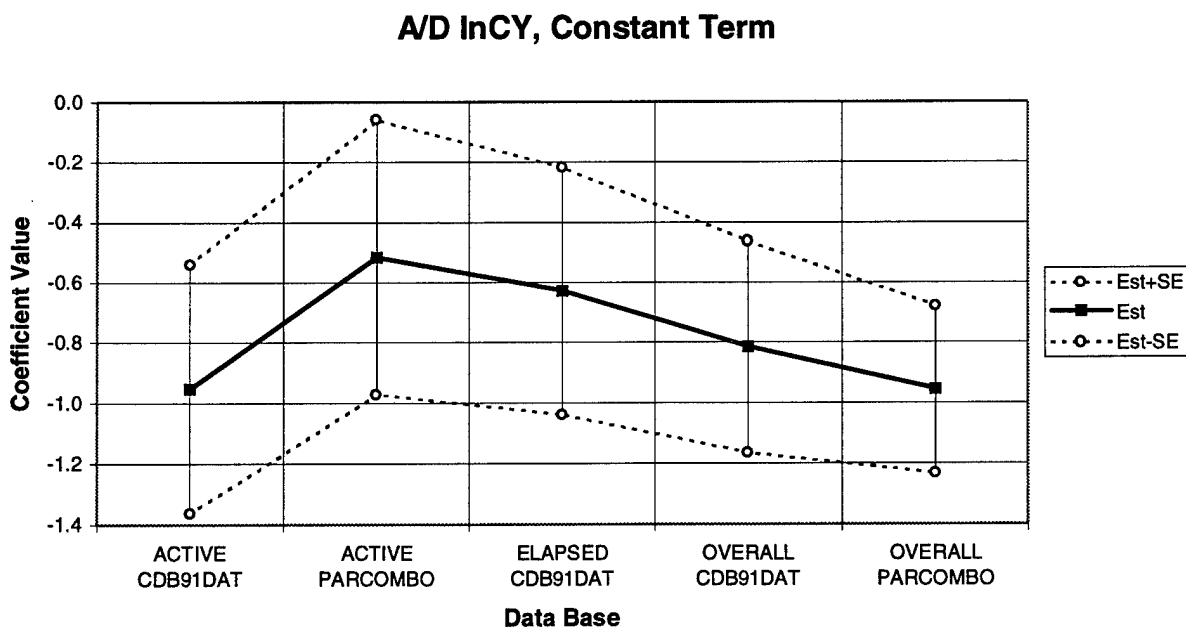
Figure 4-6. Constant Term for Regression of lnCY using Various Data Bases

A/D lnCY, Coefficients of lnX0 and lnY0
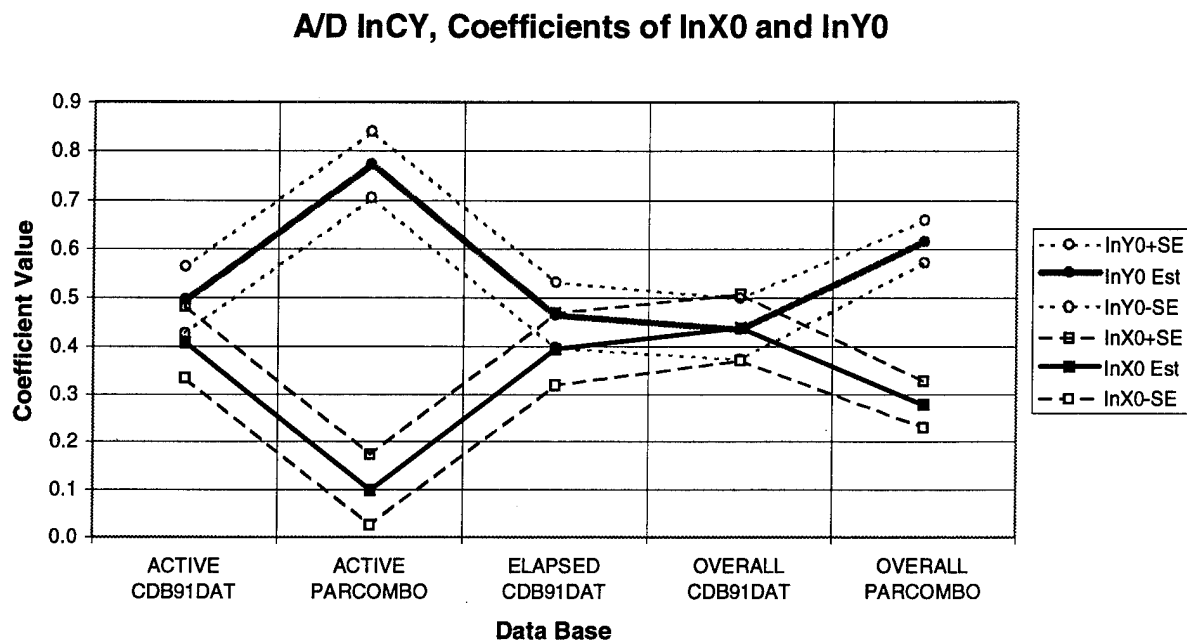
Figure 4-7. Coefficients of lnX0 and lnY0 for Regression of lnCY using Various Data Bases

A/D lnCY, Coefficient of T
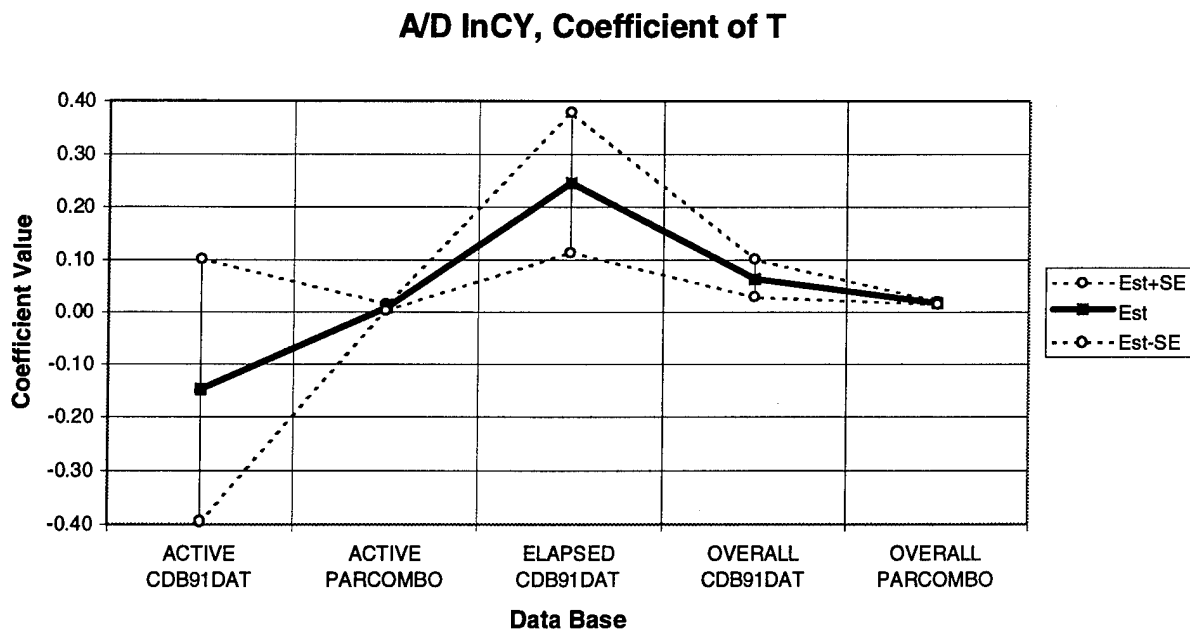
Figure 4-8. Coefficient of T for Regression of lnCY using Various Data Bases

A/D lnCY, Coefficient of D
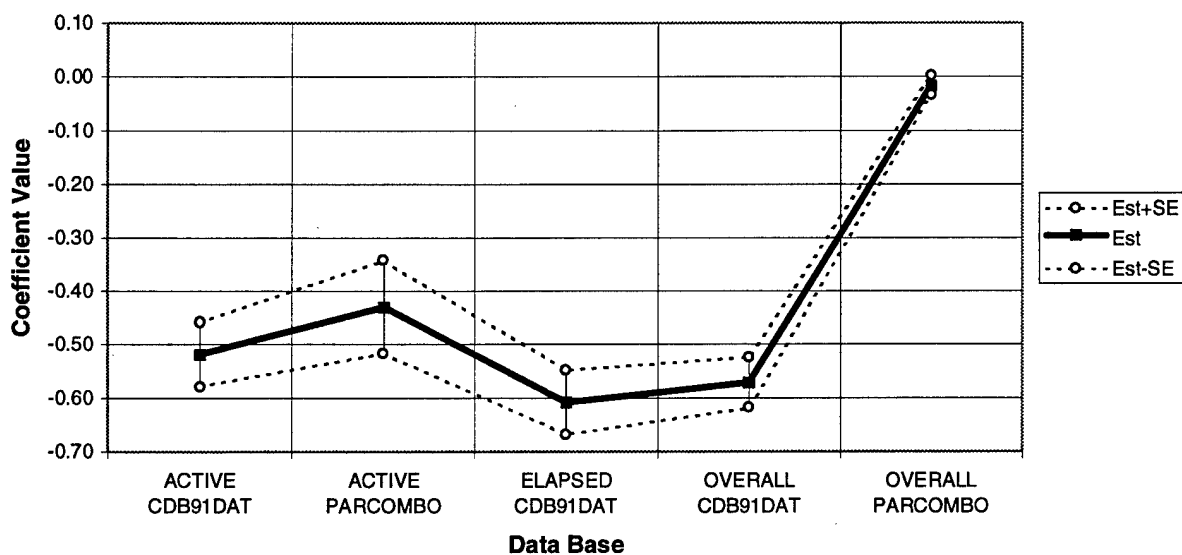
Figure 4-9. Coefficient of D for Regression of lnCY using Various Data Bases

A/D lnCERY, Constant Term
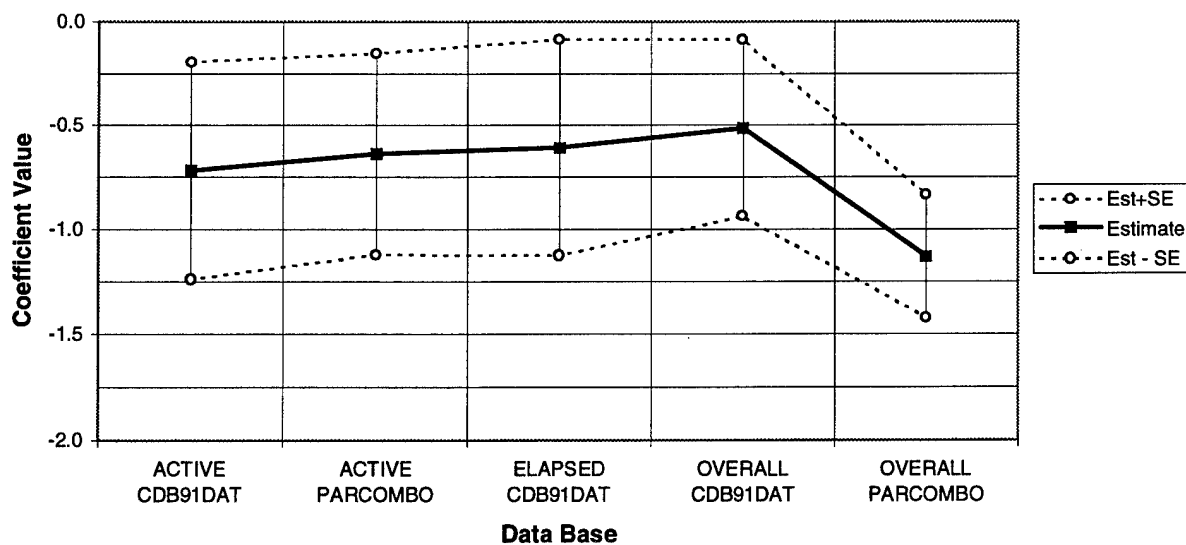
Figure 4-10. Constant Term for Regression of lnCERY using Various Data Bases
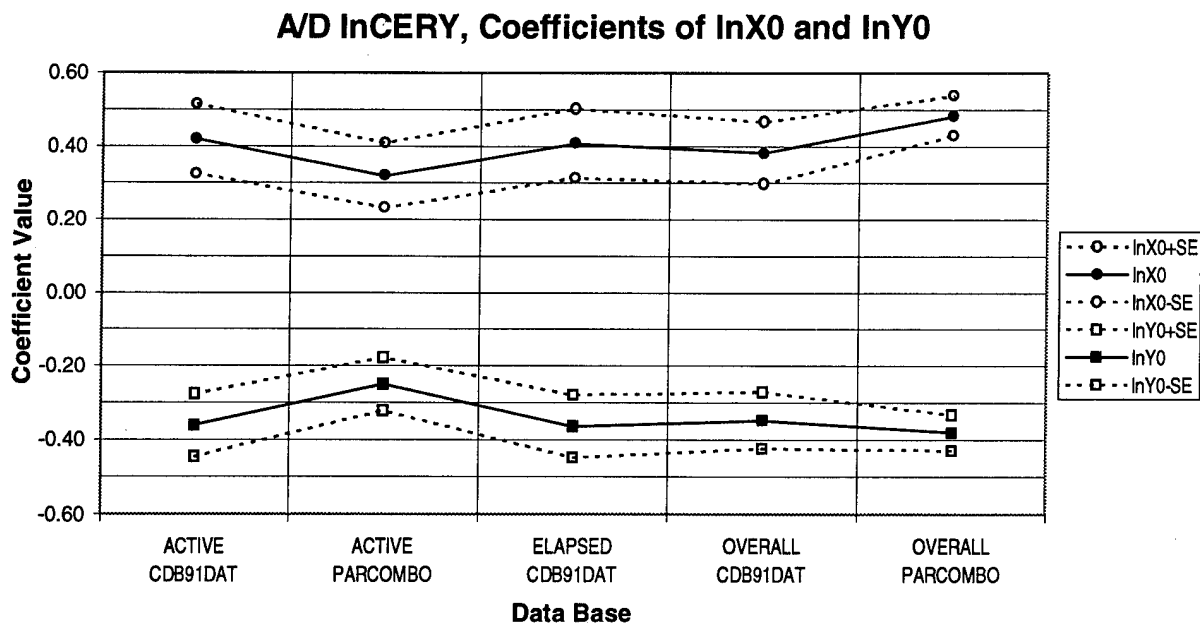
Figure 4-11. Coefficients of lnX0 and lnY0 for Regression of lnCERY using Various Data Bases
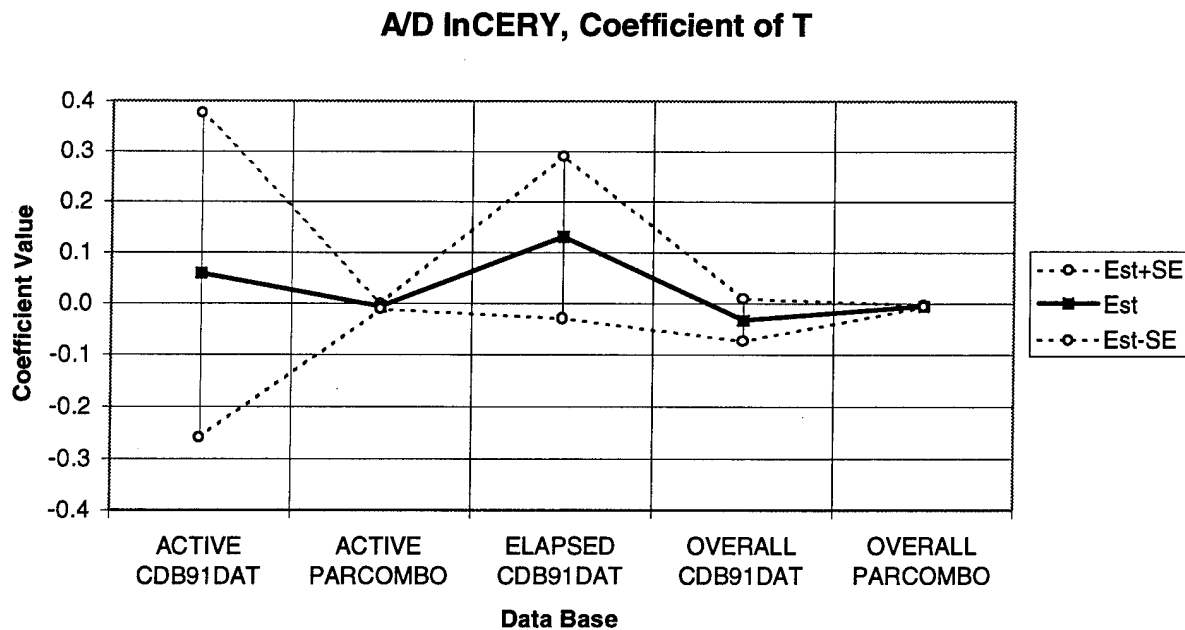
Figure 4-12. Coefficient of T for Regression of lnCERY using Various Data Bases

A/D lnCERY, Coefficient of D
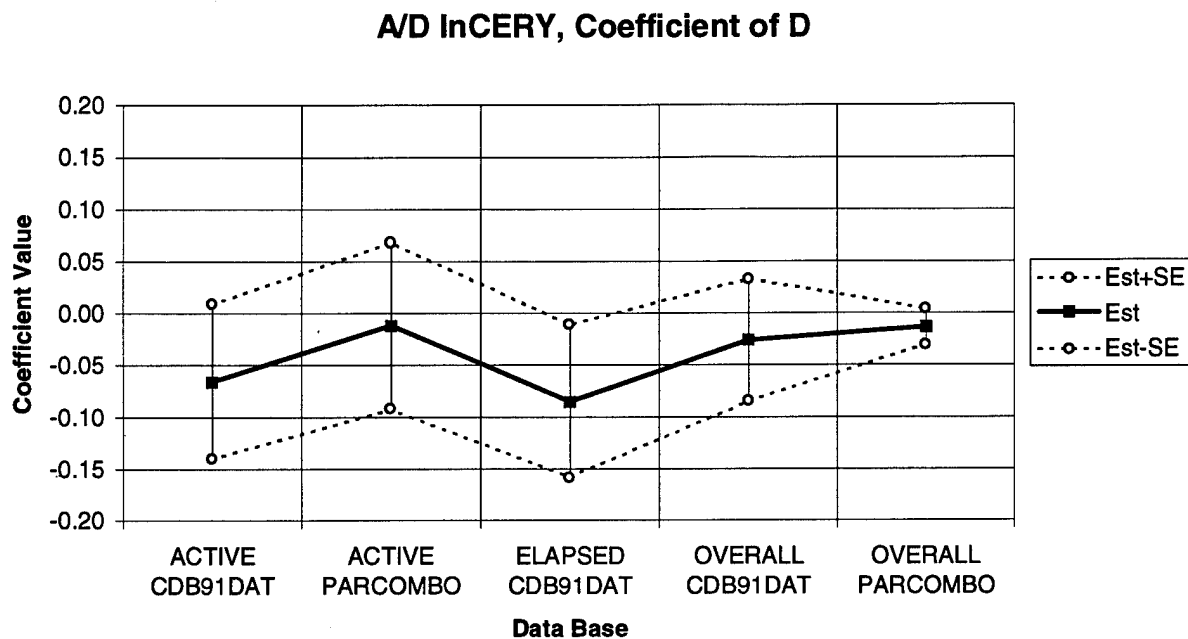
Figure 4-13. Coefficient of D for Regression of lnCERY using Various Data Bases

A/D lnCERY, Sum of Coefficients of lnX0 and lnY0
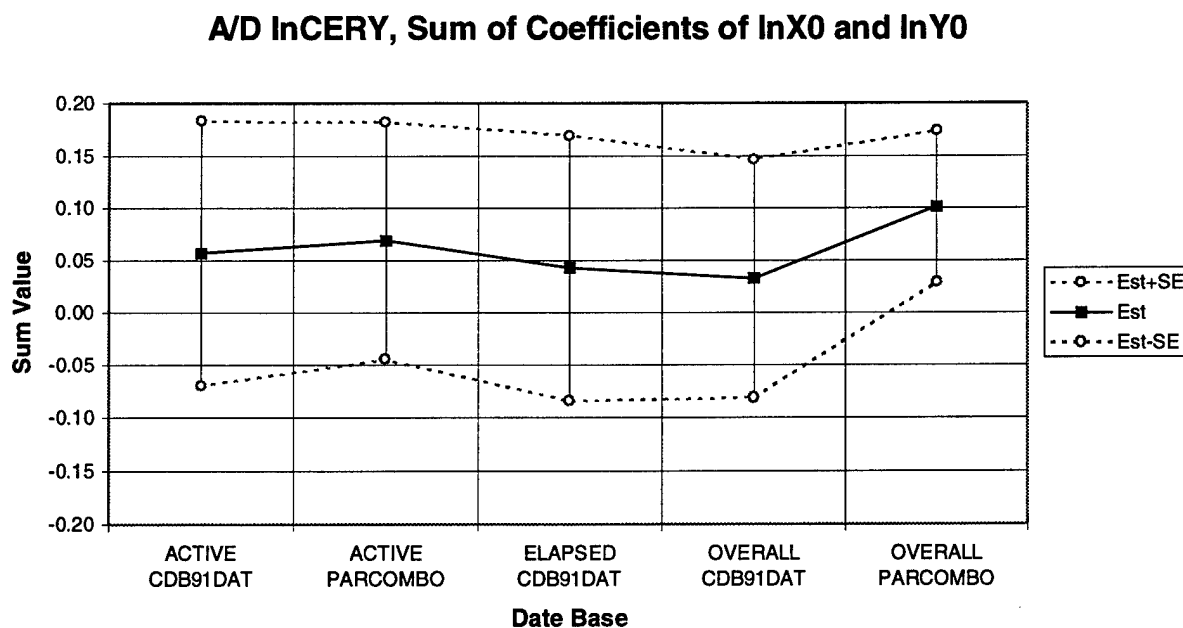
Figure 4-14. Sum of lnX0 and lnY0 coefficients for Regression of lnCERY using Various Data Bases

W/L lnCW, Constant Term
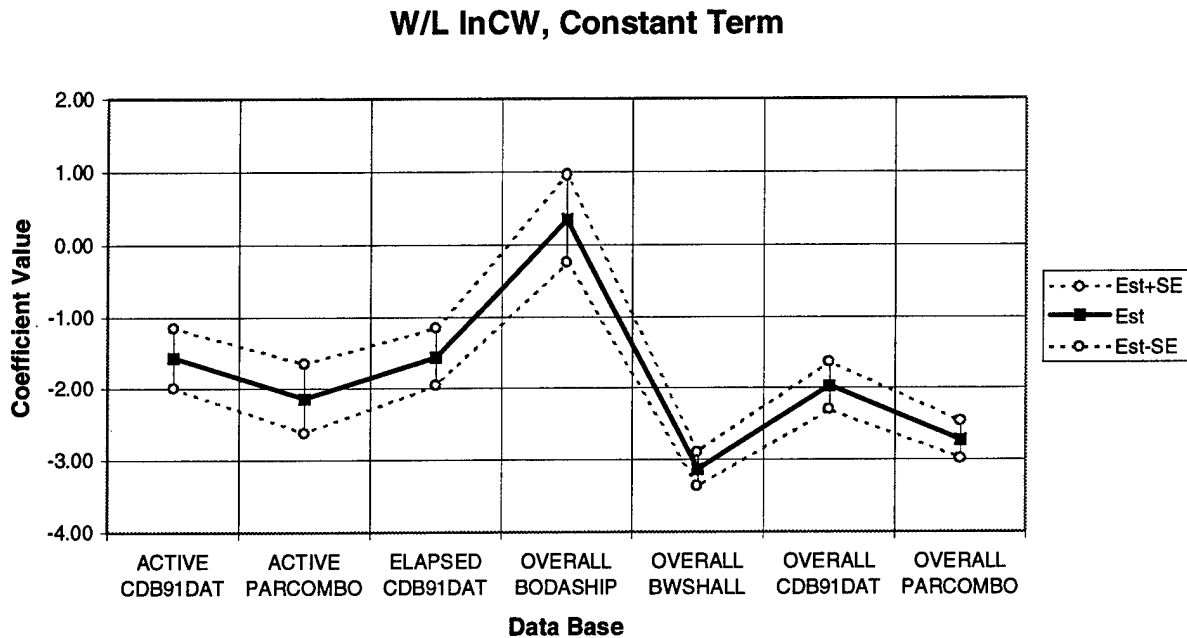
Figure 4-15. Constant Term for Regression of lnCW using Various Data Bases

W/L lnCW, Coefficients of lnW0 and lnL0
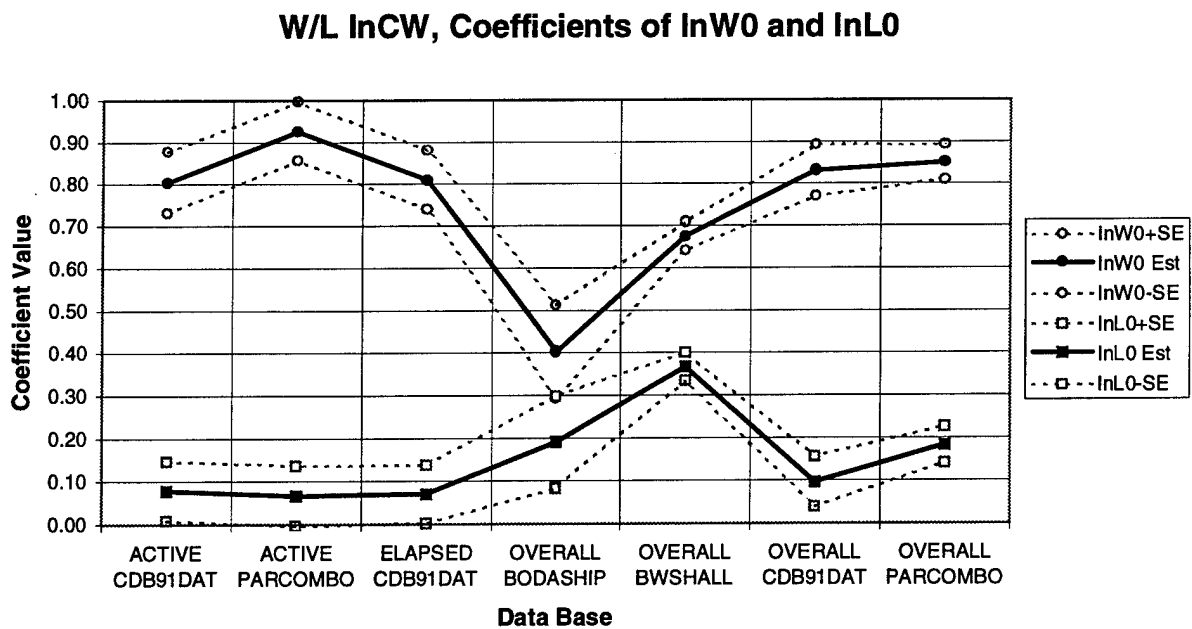
Figure 4-16. Coefficients of lnW0 and lnL0 for Regression of lnCW using Various Data Bases

W/L lnCW, Coefficient of T
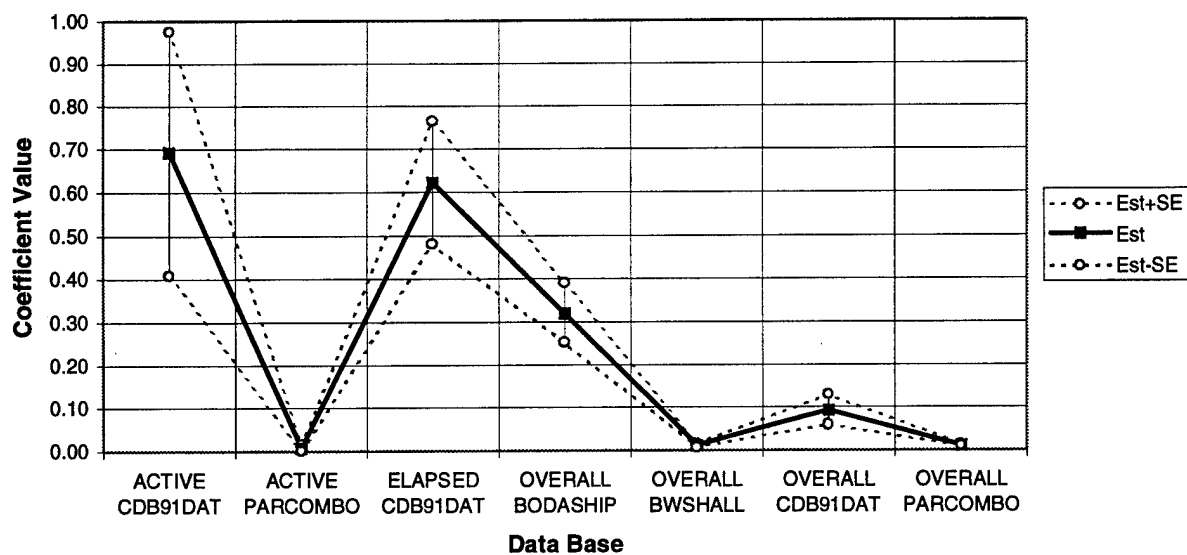
Figure 4-17. Coefficient of T for Regression of lnCW using Various Data Bases

W/L lnCW, Coefficient of D
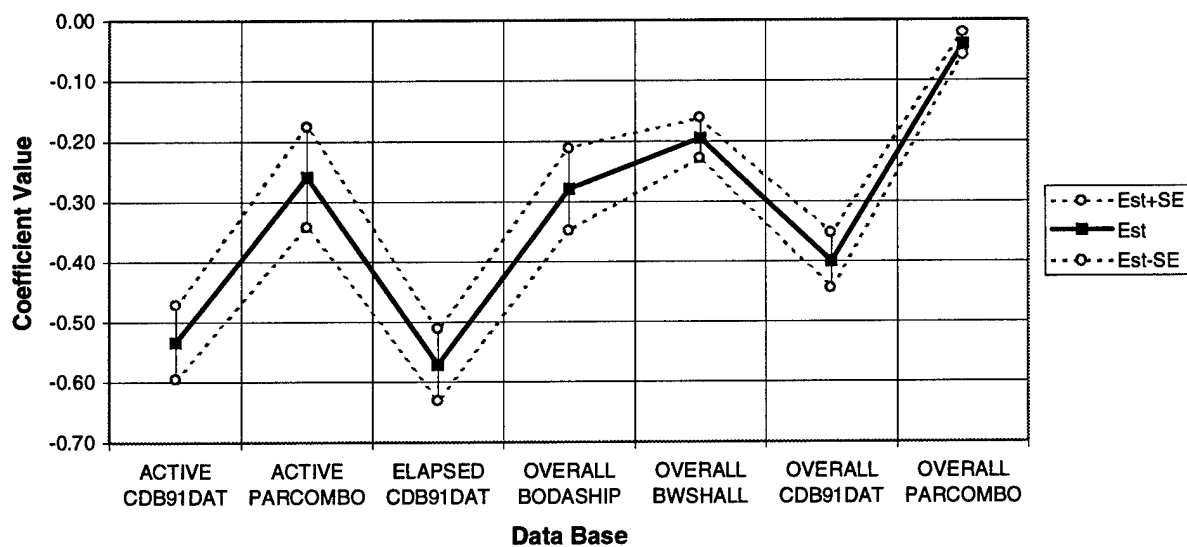
Figure 4-18. Coefficient of D for Regression of lnCW using Various Data Bases

CAA-RP-95-1

W/L lnCL, Constant Term
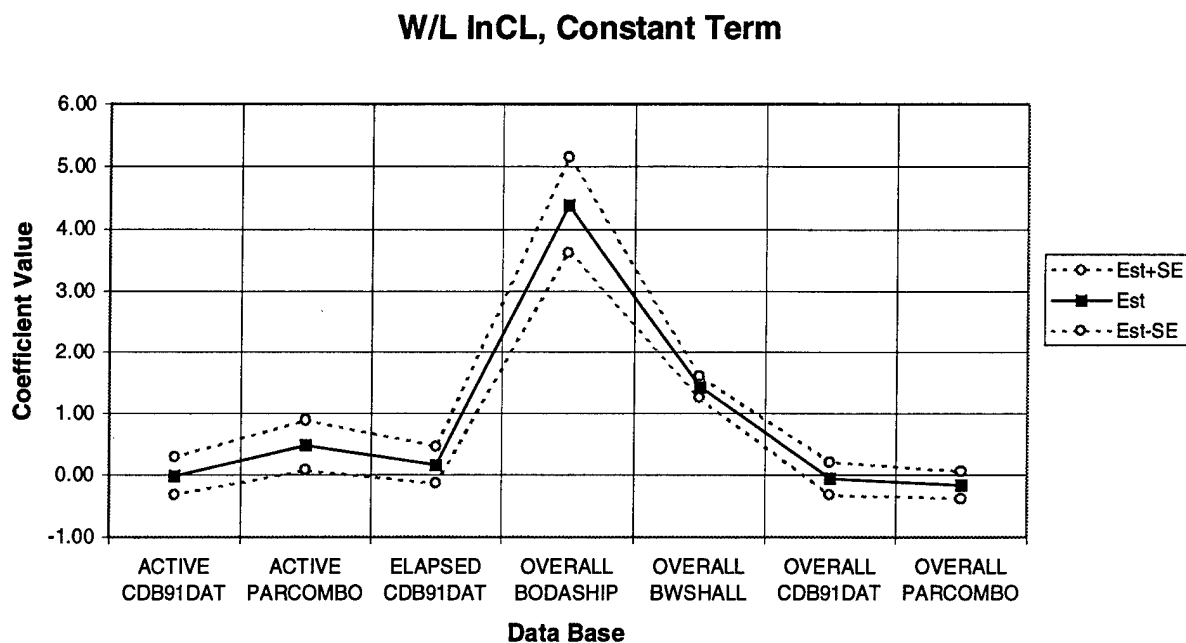
Figure 4-19. Constant Term for Regression of lnCL using Various Data Bases

W/L lnCL, Coeffcients of lnW0 and lnL0
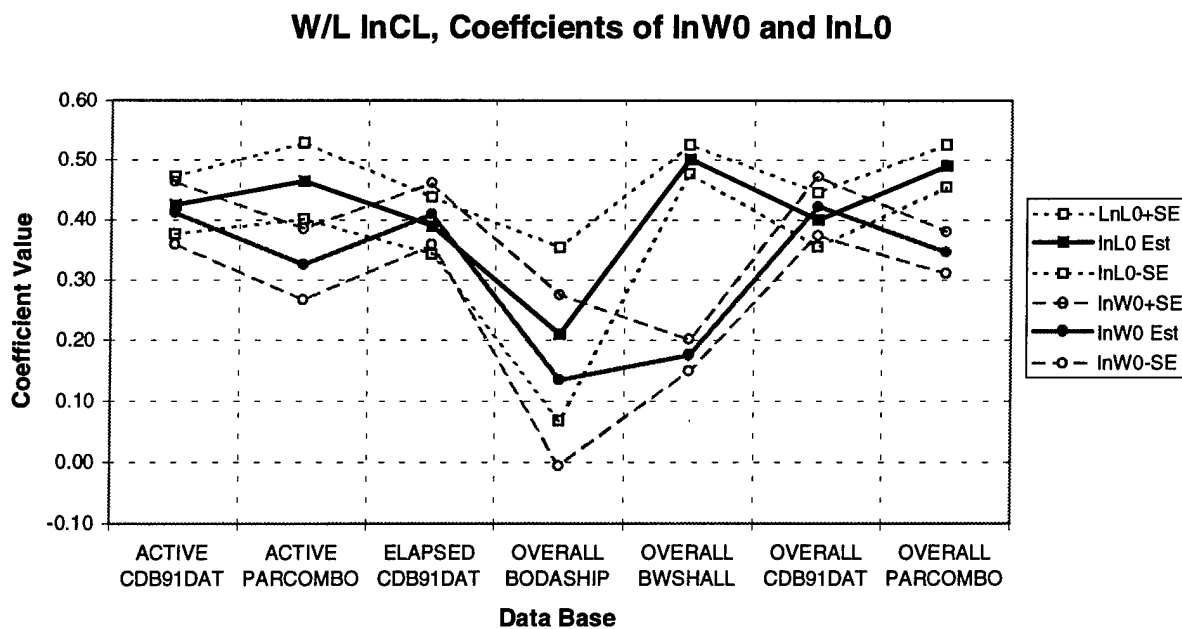
Figure 4-20. Coefficients of lnW0 and lnL0 for Regression of lnCL using Various Data Bases

4-42

W/L lnCL, Coefficient of T
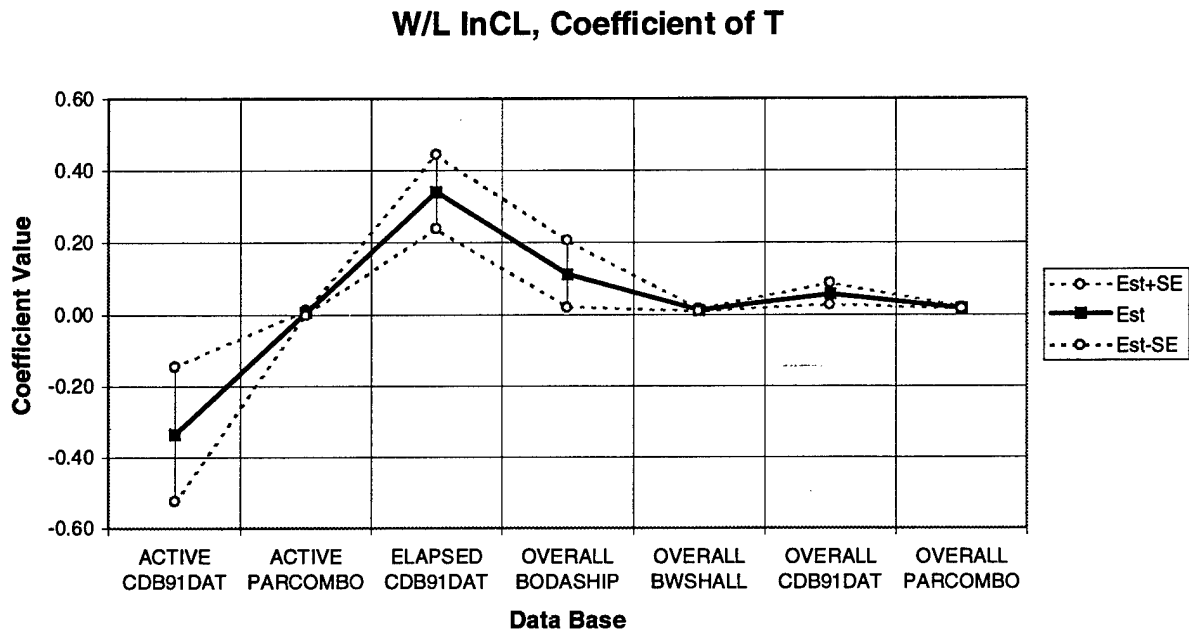
Figure 4-21. Coefficient of T for Regression of lnCL using Various Data Bases

W/L lnCL, Coefficient of D
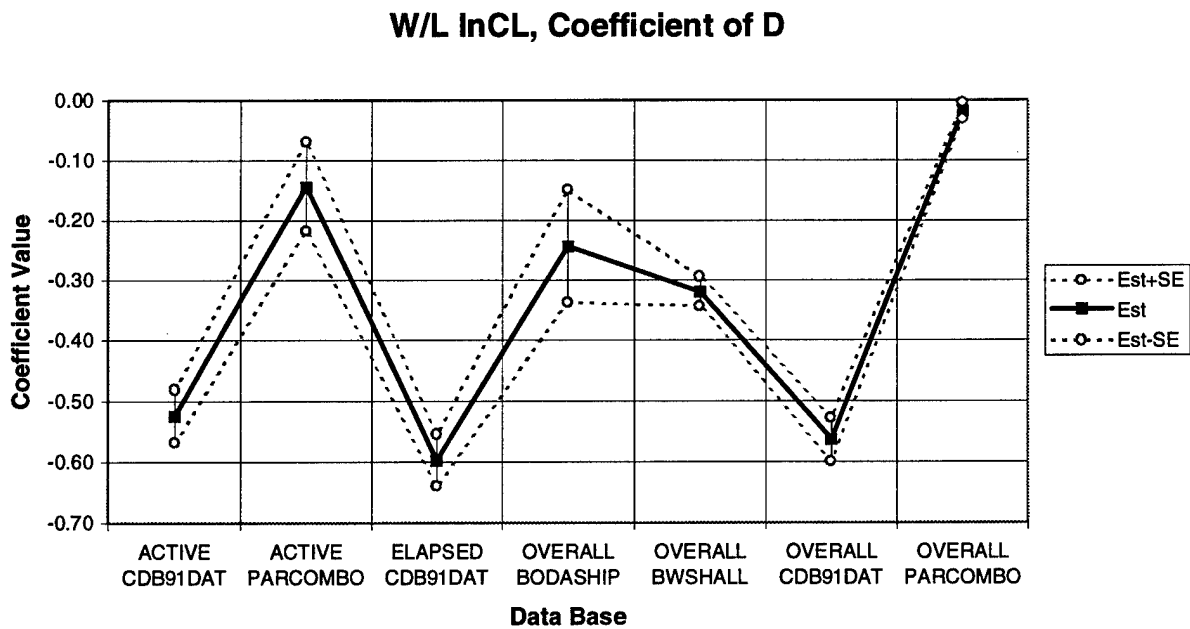
Figure 4-22. Coefficient of D for Regression of lnCL using Various Data Bases

A/D lnCERW, Constant Term
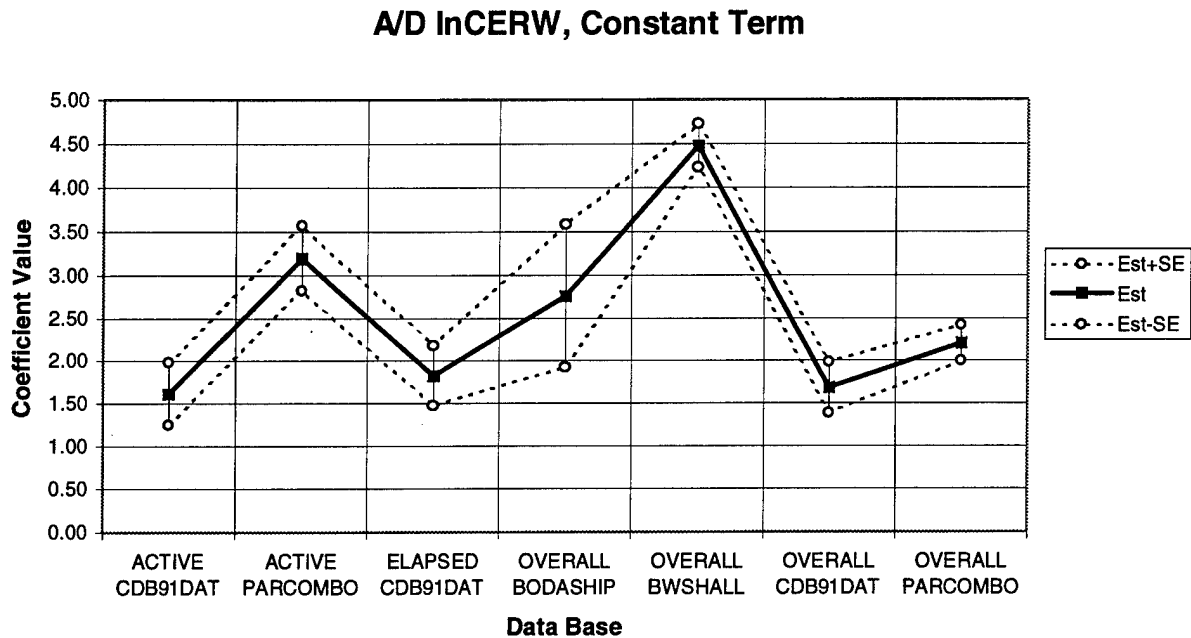
Figure 4-23. Constant Term for Regression of lnCERW using Various Data Bases

W/L lnCERW, Coefficients of lnW0 and lnL0
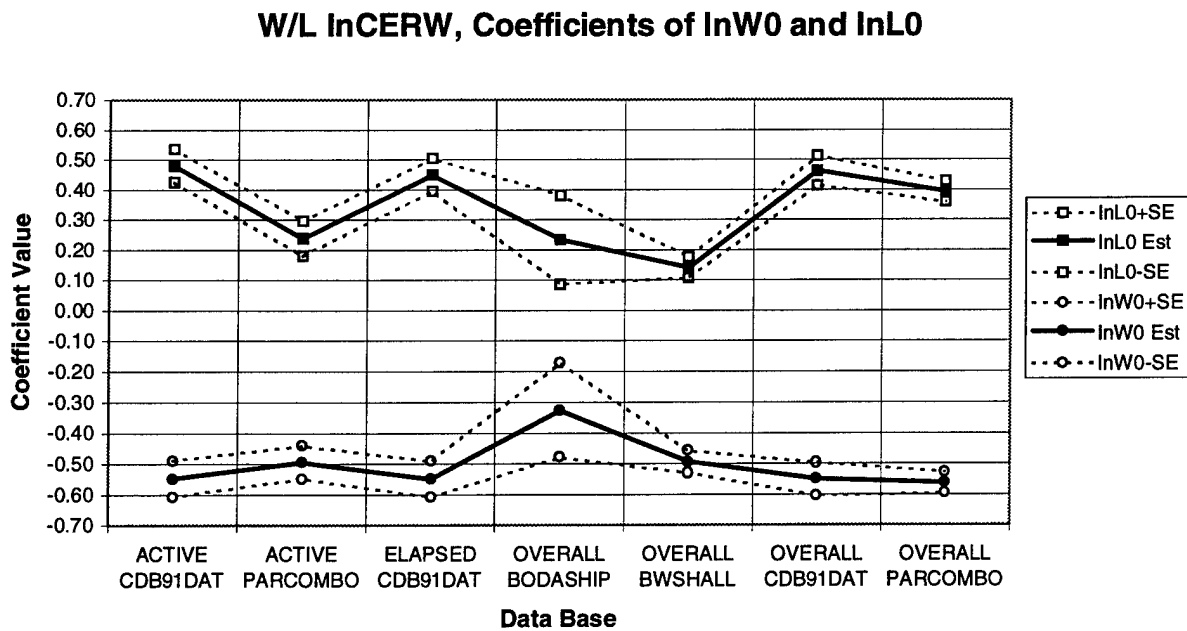
Figure 4-24. Coefficients of lnW0 and lnL0 for Regression of lnCERW using Various
Data Bases

W/L CERW, Coefficient of T
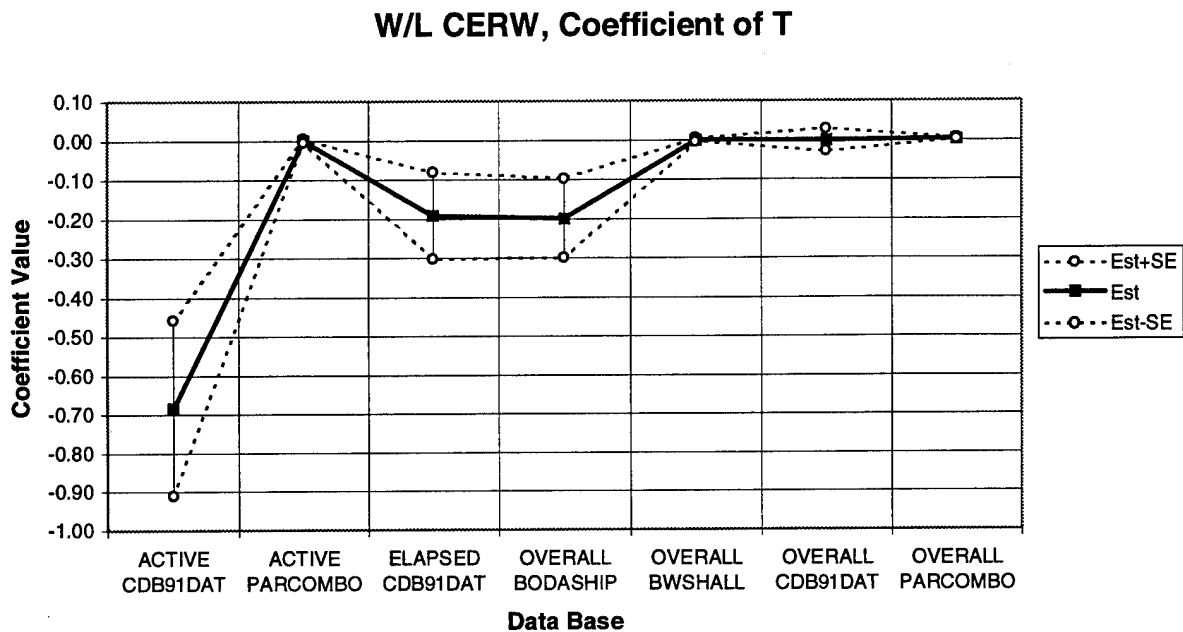
Figure 4-25. Coefficient of T for Regression of lnCERW using Various Data Bases

W/L CERW, Coefficient of D
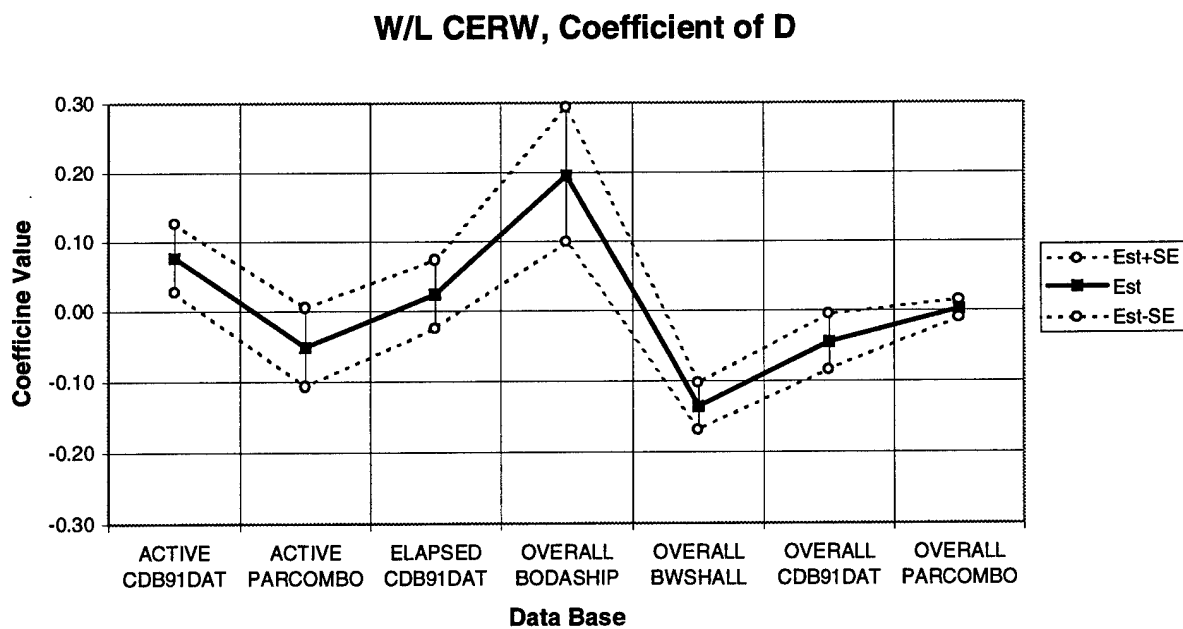
Figure 4-26. Coefficient of D for Regression of lnCERW using Various Data Bases
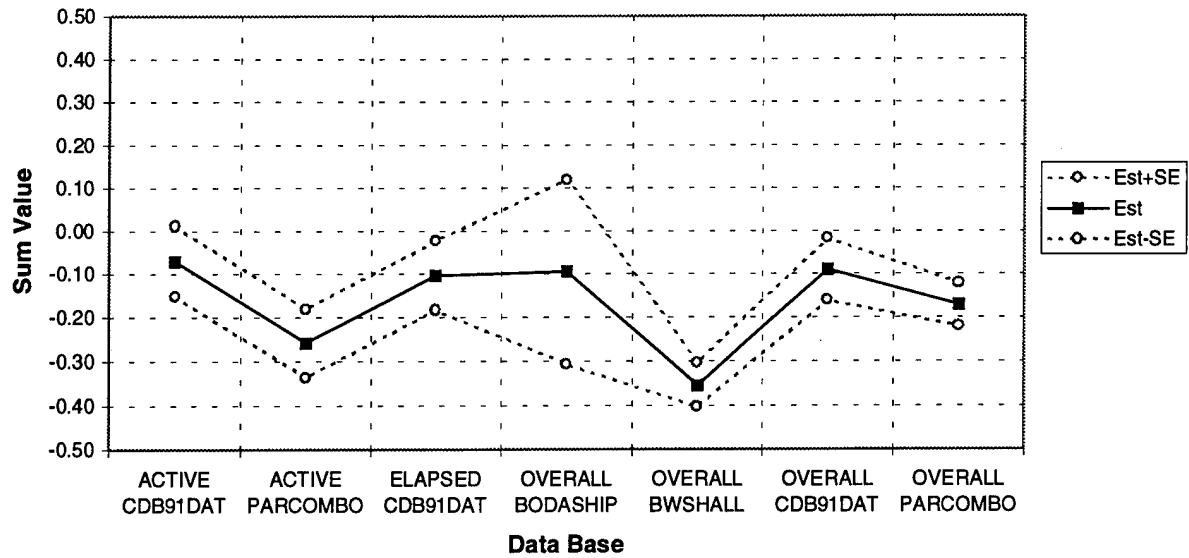
Figure 4-27. Sum of the lnW0 and lnL0 coefficients for Regression of lnCERW using Various Data Bases

CHAPTER 5

SUMMARY OF FINDINGS, OBSERVATIONS, AND CONCLUSIONS

**5-1. INTRODUCTION.** Here we summarize the findings, observations, and conclusions of this work. For the notation used, see the Glossary, paragraph 2-2, and paragraph 4-2. One general observation is that the data are highly variable, so that trends and tendencies are not always adequate summaries of the data. For an appreciation of this, the reader is referred to the figures accompanying Chapters 3 and 4. We remind the reader that the basic approach used in this study was to review the prior work in this area and then to analyze the available data bases for information related to long-term trends in personnel attrition. We focused on the analysis of the general trends in and relations among force sizes, battle durations, and casualties. Our efforts seek to advance the state of the art over prior efforts by (i) giving the Constant Fallacy (Helmbold-1994) appropriate recognition, (ii) using a regression model that includes the battle duration and battle date as potentially important factors, (iii) employing robust regression to minimize the distorting effects of a few gross errors in the data, (iv) systematically using more than one data base at a time in order to determine the sensitivity of the results to different sets of data, and (v) using several dependent variables, to include the casualty numbers as well as the casualty exchange ratio. The primary data analysis technique used is descriptive statistics.

5-2. SUMMARY OF FINDINGS REGARDING LONG-TERM TRENDS WITH RESPECT TO BATTLE DATE

    a. Battle durations have tended to increase.

    b. Attacker and defender strengths have been fairly stable over time and tended to be nearly equal. The force ratio favoring the defender has been fairly stable over time, and defenders typically fought at a slight numerical disadvantage.

    c. Attacker and defender TBC casualty numbers have declined over time and tended to be nearly equal. The casualty exchange ratio favoring the defender appears to have been fairly stable and close to unity.

    d. Attacker and defender TBC casualty fractions have declined over time, and the defender's TBC fraction has tended to be greater than the attacker's. The fractional exchange ratio favoring the defender has been relatively stable over time.

    e. Winner and loser strengths exhibit different trends with different data bases, some data bases showing an increase and others either a decrease or no appreciable change. However, all the data bases agree that the force ratio favoring the winner has been stable and close to unity.

    f. Winner and loser TBC casualty numbers have declined over time, with the loser's casualties typically at least twice those of the winner. The casualty exchange ratio favoring the winner has been more or less stable over time, depending on the data base used.

g. Winner and loser TBC casualty fractions have declined over time, with the loser's casualty fraction typically at least twice that of the winner. The fractional exchange ratio favoring the winner has been fairly stable over time.

h. Some of the trends differ from one data base to another. However, all of the data bases agree that strength is not particularly associated with victory in battle, and that both the casualty and the fractional exchange ratio are incomparably more strongly associated with victory in battle. A more refined analysis of the matter (Helmbold-1986), has shown that the fractional exchange ratio is a decidedly better index of victory in battle than the casualty exchange ratio.

5-3. SUMMARY OF FINDINGS REGARDING THE GENERAL RELATIONSHIPS AMONG FORCE SIZES, BATTLE DURATIONS, AND CASUALTIES

a. It is not true that the defender has some inherent advantages over the attacker. In fact, the attacker has generally taken fewer TBC than the defender. On the average, the casualty exchange ratio favoring the defender is less than 1, and the fractional exchange ratio favoring the defender is also less than 1.

b. Smaller forces take and inflict proportionately more casualties than larger forces. It is conjectured that this is a result of diminishing returns to scale.

c. Neither Lanchester's square law, Osipov's law, nor Peterson's logarithmic law are good approximations to the true relation of casualty fractions to force ratios.

d. The data do not reveal the expected (that is, approximately linear) dependency of casualty numbers on the temporal duration of a battle. One might expect that the casualty numbers would be in direct proportion to the duration of a battle, but this is clearly not what the data show.

e. The casualty exchange ratio favoring the defender *decreases* as the force ratio favoring the defender *increases*, despite what one might have expected. In fact, an approximate equation relating *CERY* to *FRY* is

$$CERY \approx \frac{1}{2}(FRY)^{-0.4}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

f. However, the fractional exchange ratio favoring the defender does increase as the force ratio favoring the defender increases. An approximate equation relating *FERY* to *FRY* is

$$FERY \approx \frac{1}{2}(FRY)^{+0.6}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

g. The casualty exchange ratio favoring the winner *decreases* as the force ratio favoring the winner *increases*, despite what one might have expected. In fact, an approximate equation relating CERW to FRW is

$$CERW \approx 7.4(FRW)^{-0.5}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

h. However, the fractional exchange ratio favoring the winner does increase as the fore ratio favoring the winner increases. An approximate equation relating *FERW* to *FRY* is

$$FERW \approx 7.4(FRW)^{+0.5}.$$

As indicated, the battle duration and battle date appear to have very little influence on this relation.

i. Other casualty relations appear to differ, depending on the data base use, the battle duration, and the battle date. The reasons for these differences are left to future investigations.

APPENDIX A

CONTRIBUTORS

A-1. TEAM

a. Research Director

Dr. Robert L. Helmbold, Tactical Analysis Division

A-2. PRODUCT REVIEW BOARD

Mr. Ronald J. Iekel, Chairman

Dr. Yuan-Yan Chen

Mr. Gerald E. Cooper

APPENDIX B

STUDY DIRECTIVE

CSCA-MV

MEMORANDUM FOR SPECIAL ASSISTANT FOR MODEL VALIDATION

SUBJECT: Personnel Attrition Rates in Land Combat Operations, Phase 2 (PAR-P2)

1. PURPOSE OF THE STUDY DIRECTIVE. This Directive provides tasking and guidance for the conduct of the Personnel Attrition Rates in Land Combat Operations, Phase 2 (PAR-P2) study effort, which will prepare selected computerized databases on personnel attrition rates in land combat operations and compute some basic descriptive statistics using them.

2. BACKGROUND.

   a. The results of US Army models and war games of combat are continually being challenged to demonstrate their validity. One of the key features of military combat is the infliction and suffering of personnel attrition. To provide an adequate basis for assessing the validity of US Army war games and models of combat, it is necessary that the reported data and past studies of personnel attrition rates in actual combat operations be (i) collected, (ii) reviewed and critiqued to highlight their salient characteristics, (iii) organized in a systematic form for convenient analysis, and (iv) that the most important data be subjected to an independent analysis.

   b. PAR as a whole is conceived as including three broad phases, as follows:

      (1) Phase 1. Obtain and prepare a comprehensive bibliography of the available data and past studies on personnel strengths and attrition rates in land combat operations.

      (2) Phase 2. Survey and review the data and past studies, and put the data into readily analyzable electronic form.

      (3) Phase 3. Perform some original analyses of the assembled data.

   c. PAR is limited to studying personnel strengths and battle casualties of land combat forces. Other types of attrition (nonbattle losses, losses to equipment, casualties to other services, and so forth) are outside PAR's scope. PAR is concerned only with historical data on actual combat operations; it will not deal with personnel losses in models, simulations, war games, field experiments, or training exercises (like those of the National Training Center). PAR will focus mainly on either original or translated works in English, although the most important works in other languages should be included. Studies of personnel attrition are also included, provided they contain cogent analyses of a publicly available, nonproprietary body of tabulated data on attrition in actual combat operations. Since trends in attrition over long periods of time are of interest, data on ancient as well as recent battles are solicited. However, as no contract

support is anticipated and in-house resources are limited, no systematic effort will be made to extract data from the archives or primary source materials, and no original historical research is envisioned. Thus, PAR will rely almost exclusively on secondary works that contain data in readily usable tabulated form.

d. Phase 1, or PAR-P1, was devoted to assembling the available data and past studies on personnel strengths and attrition rates in land combat operations, preparing a comprehensive annotated bibliography of it, and planning the approach to subsequent phases. It provided an annotated bibliography of over 200 relevant works, with several different types of indexes to aid retrieval.

e. One of the requirements of PAR-P1 was to plan for subsequent phases of PAR. In this regard, PAR-P1 listed several candidate hypotheses for consideration in subsequent phases of the PAR studies, and a version of that list is included at ENCL 1 for ease of reference.

3. STUDY SPONSOR AND SPONSOR'S STUDY DIRECTOR. The Director, US Army Concepts Analysis Agency (CAA) will sponsor this study. The Sponsor's Study Director will be Dr. Robert L. Helmbold of the Office of the Special Assistant for Model Validation (SAMV).

4. STUDY AGENCY. CAA's Scenarios and Model Validation Division will conduct this study. Augmentation and assistance will be provided as outlined in Paragraph 6 of this Study Directive.

5. TERMS OF REFERENCE.

a. Scope. This study directive is intended to provide for PAR-P2, the second phase of the Personnel Attrition Rates (PAR) study.

b. Objectives. The main objectives of PAR-P2 are to (i) publish a CAA Research Paper on the combat data bases currently available in digital form for use on personal computers, (ii) compute from those databases some basic statistics and publish a CAA Research Paper describing them, and (iii) plan for subsequent phases of the PAR study.

(1) A major objective of PAR-P2 is to publish a CAA Research Paper on the combat databases currently available. A coordinate objective is to provide diskette copies of these databases to the Defense Technical Information Center (DTIC) for archival storage. The criteria for inclusion of a database are as follows (roughly in order of importance). The database must be:

(a) In the public domain, so that copies can be made available to Governmental agencies and others without restriction and for (at worst) a nominal cost. However, for the sake of completeness, some important proprietary databases can be described, even if their data cannot be made available through DTIC.

(b) In data base form (i.e., consist primarily of tabulations rather than narratives).

(c) Such as to contain information on military operations in and/or outcomes of battles or wars.

(d) Available on diskettes usable with personal computers. (Some of the databases in document form collected during PAR-P1 may be converted to digital form during PAR-P2 if that appears to offer a significant benefit to subsequent phases of the PAR study. However, the extent of such digitization will be drastically limited by available time and effort.)

(e) Useful to many military operations analysts; developers, users, and assessors and validators of the inputs and/or outputs of war games and analogous combat simulations; military historians; students of military art and science; and others with similar interests.

(f) Difficult or inconvenient for individuals and separate study teams to generate or recreate, but which would be used frequently if readily available through DTIC.

(2) The combat databases are envisioned to include at least those listed below. In some cases, we may find that copyright or other restrictions prevent inclusion of the actual data. In such cases, we will describe the database and its availability, but will not make the actual data available for general use.

- ACSDB-1990, Ardennes Campaign Simulation Data Base (ACSDB).
- AHART-1993, Analysis of Historical Artillery Expenditures (AHART) Study CY87.
- BENCHMARK-1988, Data on Historical Battles 1937-1982 (Benchmarks).
- BERNDT, Data from Berndt's *Zahl im Kriege*.
- BOB18, Data on the air warfare that took place during the Battle of Britain.
- BWSH-1993, Bodart-Willard-Schmieman-Helmbold data base of 1,087 battles.
- CDB90-1991, CAA's Database of Battles--Version 1990.
- CRETE, CNA's database of Crete.
- CREWCAS, Visco's tank crew casualty data base.
- INCHON, Busse's data on the Inchon-Seoul campaign.
- IWOJIMA, Various interpretations of the Iwo Jima casualty experience.
- KELLEY-1977, Air combat engagement database.
- LIVERMORE, The Livermore data on US Civil War battles.
- Logistics Management Institute database of Twelfth Army casualty experience, collected by George Kuhn.
- POGOGORO, Data on the Pogoroloye-Gorodische battle.
- ROADATA-1991, ROADATA Rates of advance Data.
- SINGER, Extracts from Singer's data on wars.
- SMALL, Extracts from Small's data on wars.
- SP128-1961, Historical data and Lanchester's theory of combat.
- SP190-1964, Historical data and Lanchester's theory of combat: Part II.
- WESTWALL, Data on the Westwall battle of World War II.

(3) Each database will be described using the following categories:

(a) General (full bibliographic reference to the primary documentation, description of the kinds of information it provides, other important factual information, and a statement of the situational descriptors used).

(b) Data Sources Used.

(c) Diskette Format (computer hardware and software compatibility restrictions, file descriptions, data field specifications, and so forth).

(d) Other and Miscellaneous (examples of the use of this database, other informative remarks).

(e) Comments and Critique (discussion of the strong and weak points of this database).

(4) Compute some basic descriptive statistics using these databases and publish a separate CAA Research Paper documenting them. These will be selected from the list at ENCL 1. However, it is desired that they include at least the following:

(a) Distribution of daily values for a given unit.

(b) Distribution among different units for a given day.

(c) Joint distribution by units and days.

(d) Ratio of KIA to WIA over time.

(e) Distribution of hits on different parts of the body over time.

(f) Attrition rates over the course of time for battles of the last 400 years.

(5) Plans for the conduct of Phase 3 of PAR will be developed during PAR-P2.

c. Timeframe. Not applicable.

d. Assumptions. Not applicable.

e. Essential Elements of Analysis for PAR-P2.

(1) What computerized databases are or can be made available to support research in personnel attrition rates during subsequent phases of the PAR work?

(2) What research topics will these materials support?

(3) What would be an efficient way to conduct such research?

f. Environmental and Threat Guidance. Not applicable.

g. Estimated Cost Savings or Other Benefits.

(1) It is important that the validity (or range of validity) of US Army war games and models of combat be assessed as accurately as possible. This can only be done through the application of the scientific method to historical data. This study is a necessary step in that process.

(2) US Army studies and analyses often need summary quantitative relationships applicable throughout a broad range of combat situations. It would be costly and inefficient to have each study review the literature, assemble the applicable information, convert it to electronic form, and make its own analyses of the reported data on personnel attrition. Making the results of this study available to a wide audience will help avoid unnecessary duplication of effort.

6. RESPONSIBILITIES. CAA's Scenarios and Model Validation Division will conduct the study. Administrative support will be provided by CAA-MS.

7. LITERATURE SEARCH. A detailed annotated bibliography of sources was prepared during PAR-P1. While no formal literature search is specifically planned for subsequent phases of the PAR studies, we intend to continue informal efforts to identify and acquire additional relevant data.

8. REFERENCES.

a. Administrative and Procedural.

b. Substantive. "Personnel Attrition Rates in Land Combat Operations: An Annotated Bibliography," US Army Concepts Analysis Agency Research Paper, CAA-RP-93-2, August 1993.

9. ADMINISTRATION.

a. Funding. Funding will be provided by CAA.

b. Administrative Support. Administrative support will be provided by CSCA-MS.

c. Cost Limitations. Not applicable.

d. Contract Studies. Not applicable.

e. Automatic Data Processing Equipment (ADPE) Support. Personal computers and associated equipment (such as monitors, printers, etc.) will be required, as will appropriate software systems for databases, spreadsheets, word processing, statistical analyses, and programming languages such as BASIC. No need is currently anticipated for other ADPE support.

f. Milestone Schedule. The published Research Papers describing the available combat databases and documenting the basic statistics, together with the draft Study Directive for Phase 3 and its supporting ARB presentation, are to be completed by 31 December 1994.

g. Sponsor's Study Director (SSD) & Study Advisory Group (SAG). Not applicable.

h. Responsibility for DD Form 1498. Scenarios and Model Validation Division, MV.

i. Study Format. The catalog of available databases and the basic statistical analyses are to be documented as separate CAA Research Papers. An outline approach to subsequent phases is to be presented as a draft Study Directive and supporting ARB.

j. Action Documents. Written evaluation of study results will be provided by the sponsor in accord with AR 5-5.

ENCL

E. B. VANDIVER III
Director

CANDIDATE RESEARCH TOPICS FOR PAR-P2

1. Basic Descriptive Statistics.

    a. Distributions of values (such as casualties, casualty rates, casualty fractions, attrition rates, etc.).

        (1) Daily values for a given unit are distributed (normally, exponentially, log-normally, Weibull, Gamma, etc.)

        (2) Daily values are distributed among units according to a (multinomial, hypergeometric, Pareto, Maxwellian, Bose-Einstein, Fermi-Dirac, etc.) distribution.

        (3) Values are distributed over units and over days as a bivariate distribution (of what type?).

        (4) Average World War II division engagement casualty rates in Western Europe were 1% to 3% per day.

    b. Materiel loss rates are related to personnel casualty rates.

        (1) Tank loss rates are five to seven times higher than personnel casualty rates.

        (2) Attacker tank loss rates are generally higher than defender tank loss rates. (This is in relation to personnel casualty rates on the opposing sides. If the attacker's tank loss rate is about seven times that of the attacking personnel casualty rate, the defender's tank loss rate will probably be closer to five times (or even less) the defender's [personnel] casualty rate.)

        (3) Artillery materiel loss rates are generally about one-tenth personnel casualty rates.

        (4) Self-propelled artillery loss rates are two-to-three times greater than for towed guns.

        (5) The loss rates of light, to medium, to heavy artillery weapons are in the proportion: 2.2/1.8/1.0.

2. Trends (influence of the passage of years on values).

    a. Values increase as weapon lethality (effective range, rate of fire, accuracy, and terminal effectiveness) increases.

    b. Values have steadily and steeply declined with the passage of time since about 1600 AD.

    c. The distribution of hits over the body has not changed much over the course of time.

    d. The distribution of killed and wounded casualties in 20th Century warfare is constant. (About 20% of the battle casualties are killed immediately. This corresponds to a wounded to killed ratio of about 4.)

e. Values in the 1973 October War were comparable to those of World War II.

f. Values for major power forces in minor hostilities after 1945 are about half those experienced in World War II.

3. Environment (influence of environmental factors on values).

a. Values for both sides decline as the severity of the environmental conditions increases.

(1) In difficult terrain, casualty rates for both sides decline markedly.

(2) Values vary seasonally, increasing in the "good weather" seasons, and decreasing in the "bad weather" seasons. In particular, values are higher in summer than in winter.

(3) In bad weather, casualty rates for both sides decline markedly.

(4) Casualty rates are lower at night than in the daytime.

(5) Values decrease as the terrain becomes more difficult.

b. Values are largely independent of environmental factors such as weather, temperature, visibility, the degree of cover and concealment provided by the natural terrain, and like factors.

c. National characteristics.

(1) National characteristics have no large, enduring, or predictable effects on relative values.

(2) Values are strongly affected by "national characteristics."

(3) What is the effect of national character on victory and losses in battle? Some analysts claim that its influence is consistent, pervasive, and of major importance. On the other hand, some found little or no evidence for that.

4. Tactics (influence of the tactical situation on values).

a. Attack and defense.

(1) In the average modern battle, the attacker's numerical strength is about double the defender's.

(2) In most modern battles, the numerical losses of attacker and defender have been similar.

(3) In the average modern battle, attacker casualty rates are somewhat lower than defender casualty rates.

b. Values decline as speed of movement increases. The faster the front moves, the lower the casualty rates for both sides.

c. Values are "episodic" and tend to occur as "crisis waves."

(1) The casualty-inflicting capability of a force declines after each successive day in combat.

(2) Values decline with time into a battle, campaign, or war.

(3) The longer the battle (campaign, or war) the lower its values.

d. How are the values affected by various tactical factors (such as weather; terrain; fortifications; state of training, morale, unit cohesion; combat experience and so forth; nature and effectiveness of the command, control, intelligence, and communications system; type of tactical disposition or maneuver of forces; availability and effectiveness of air and fire support; and so forth)? No very satisfactory analysis of this is available.

(1) A force with greater overall combat power inflicts casualties at a greater rate than the opponent.

(2) More effective forces inflict casualties at a higher rate than less effective opponents.

(3) The side with the greater amount or quality of fire support (air and ground) has lower values relative to its opponent.

(4) Casualty rates of a surprising force are lower than those of a surprised force.

(5) Casualty rates for defenders vary inversely with strength of fortifications. (The stronger the fortifications, the lower the defender's casualty rate.)

(6) A flanking maneuver, if successful, inflicts high casualties on the opponent. But if unsuccessful, results in high casualties to the friendly side.

e. There is no direct relationship between force ratios and values.

(1) The greater the friendly/enemy force ratio, the higher the friendly/enemy casualty ratio.

(2) Small force casualty rates are higher than those of large forces. (The smaller the force, the higher it's values.)

f. Casualty rates seem to decline during river crossings.

g. An "all-out" effort by one side raises casualty rates for both sides.

5. Effects (the effects of values on the conduct or outcomes of battles and wars).

a. Attrition causes units to "break."

(1) Breakpoints occur at definite values of the casualty level, casualty rate, casualty fraction, attrition rate, casualty exchange ratio, fractional casualty ratio, force ratio, etc.

(2) Each side determines its own breakpoint (either stochastically or deterministically) based solely on its own casualty experience and without reference to what is happening on its opponent's side. The main factor determining such breakpoints is the casualty level, the casualty rate, the casualty fraction, the attrition rate, etc.

(3) Breakpoints are determined by what is happening on both sides. The main factors determining such breakpoints are the casualty exchange ratio, the casualty rate ratio, the casualty fraction ratio, the casualty fraction ratio change rate, the force ratio, the force ratio's change rate, the fractional exchange ratio, the fractional exchange ratio rate of change, etc.

b. Winning and losing.

(1) Attrition increases when the losing side has little or no opportunity to disengage.

(2) The casualties to each side depend mainly on whether it won or lost. Whether it was attacking or defending makes little difference. The attrition rate of the winner is independent of whether he was attacking or defending.

(3) In the average modern battle, the attacker is more often successful than the defender.

(4) The loser's casualty fraction is about twice the winner's. Casualty rates of winners are lower than those of losers.

(5) Values on both sides are higher when they are approximately evenly matched (as measured by the force ratio, the casualty ratio, the casualty fraction ratio, the ADV parameter, etc.) than when one side is much superior.

6. Miscellaneous (not classified elsewhere).

a. What is the correct theory of attrition in battle? It seems that the currently available data do not suffice to determine the domain of applicability to real combat of any of the innumerable equations or other models that have been proposed for attrition in battle.

b. Values are about the same on both sides.

c. Values on each side are roughly proportional to the values on the other side. Friendly and enemy casualties are directly related in the sense that they move up and down together.

d. The casualties to each side are inversely proportional to their strengths (Osipov).

e. The "personal equation" effect, well known to astronomers, also affects historical data.

APPENDIX C

REFERENCES

Berndt, Otto (Captain in the Austrian General Staff), *Die Zahl im Kriege: Statistiche Daten aus der Neueren Kriegsgeshichte in Graphischer Darstellung* [Number in War: Statistical Data from Modern Military History in Graphical Form], G. Freytag & Berndt, Vienna, 1897, 169 pp, UNCLASSIFIED, available from US Army Command and General Staff College Library (355.09 B524z).

Bodart, Gaston, *Militar-Historisches Kriegs-Lexicon (1618-1905)*, C. W. Stern, Vienna and Leipzig, 1908, 956 pp, UNCLASSIFIED, available from libraries.

Clausewitz, Carl von, *On War*, edited and translated by Michael Howard and Peter Paret, Princeton University Press, Princeton, NJ, 1976, UNCLASSIFIED, available from the pulbisher and libraries.

Dodge, Theodore Ayrault, *Caesar: A History of the Art of War Among the Romans Down to the End of the Roman Empire, With a Detailed Account of the Campaigns of Caius Julius Caesar*, The Riverside Press, Cambridge, 1900, ca. 800 pp, UNCLASSIFIED, available from libraries.

Dunn, L. P., "A Lanchester Fit to Selected Land Battles," Lulejian and Associates, February 1971, 46 pp, UNCLASSIFIED, May be available from DTIC. A copy of the original is in Helmbold's personal files. Prepared for Assistant Chief of Staff, Studies and Analysis, Headquarters, US Air Force, under Contract No. F44620-70-C-0056.

Dupuy, R. Ernest; Dupuy, Trevor N., *The Encyclopedia of Military History from 3500 BC to the Present*, Harper & Row, Publishers, New York, Library of Congress Card No. 74-81871, 1970, 1406 pp, UNCLASSIFIED, available from the publisher and libraries.

Dupuy, T. N., *Attrition: Forecasting Battle Casualties and Equipment Losses in Modern War*, HERO Books, Fairfax, VA, ISBN 0-915979-26-8, 1990, 176 pp, UNCLASSIFIED, available from publisher and libraries.

Eggenberger, David, *A Dictionary of Battles*, Thomas Y. Crowell Co., New York, Library of Congress Card No. 67-12400, 1967, 526 pp, UNCLASSIFIED, available from publisher and libraries.

Engel, Joseph H., "A Verification of Lanchester's Law," Journal of the Operations Research Society of America, Vol 2, No 2, May, 1954, pp 163-171, UNCLASSIFIED, available from libraries.

Fain, Janice B., "The Lanchester Equations and Historical Warfare: An Analysis of Sixty World War II Land Engagements," History, Numbers, and War, Vol 1, No 1 (Spring), 1977, pp 34-52, UNCLASSIFIED, available from libraries. Originally published as an unclassified paper contained in the Proceedings of the 34th Military Operations Research Symposium, December 3-5, 1974 (SECRET).

FM-101-10-1, *Staff Officer's Field Manual, Organizational, Technical, and Logistical Data*, US Army Command and General Staff College, Fort Leavenworth, KS, (vol 1 is FM 101-10-1/1, Vol 2 is FM 101-10-1/2), October 1987, (Vol 1, 1,459 pp; Vol 2, 365 pp), UNCLASSIFIED, available from Department of the Army.

Fox, William F., *Regimental Losses in the American Civil War, 1861-1865: A Treatise on the Extent and Nature of the Mortuary Losses in the Union Regiments, With Full and Exhaustive Statistics Compiled from the Official Records on File in the State Military Bureaus and at Washington*, Albany Publishing Co., Albany, NY, 1889, UNCLASSIFIED, available from libraries.

Harbottle, Thomas Benfield, *Dictionary of Battles From the Earliest Date to the Present Time*, E. P. Dutton & Co., New York, Library of Congress Card Number 66-22672, 1905 (republished 1966 by Gale Research Company, Detroit), 298 pp, UNCLASSIFIED, available from publisher and libraries.

Hartley, Dean S. III, "Confirming the Lanchestrian Linear-Logarithmic Model of Attrition," Data Systems Research & Development Program, Oak Ridge Gaseous Diffusion Plant, operated by Martin Marietta Energy Systems, Inc., for the US Department of Energy, K/DSRD-263/R1, February 1991, 136 pp, UNCLASSIFIED, available from DTIC. This is a revision to the report of the same title published in December 1990.

Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat," Combat Operations Research Group (CORG), HQ, US Continental Army Command, Staff Paper CORG-SP-128, 1 July 1961, 178 pp, UNCLASSIFIED, available from DTIC (AD-480 975). See also the companion paper Helmbold-1964a.

Helmbold, Robert L., "Historical Data and Lanchester's Theory of Combat: Part II," Combat Operations Research Group (CORG), HQ, US Combat Development Command, Staff Paper CORG-SP-190, August 1964, 132 pp, UNCLASSIFIED, available from DTIC (AD-480 109È). See also the companion paper Helmbold-1961.

Helmbold, Robert L., "Some Observations on the Use of Lanchester's Theory for Prediction," Journal of the Operations Research Society of America, Vol 12, No 5, 1964, pp 778-781, UNCLASSIFIED, available from libraries.

Helmbold, Robert L., "A Modification of Lanchester's Equations," Operations Research, Vol 13, no 5, Sep-Oct 1965, pp 857-859, UNCLASSIFIED, available from libraries.

Helmbold, Robert L., "Air Battles and Ground Battles–A Common Pattern?," The RAND Corporation, RAND P-4548, January 1971, 15 pp, UNCLASSIFIED, available from DTIC (AD-718 975).

Helmbold, Robert L., Combat History Analysis Study Effort (CHASE): Progress Report for the Period August 1984-June 1985, US Concepts Analysis Agency, US Army Concepts Analysis Agency Technical Paper CAA-TP-86-2, August 1986, 249 pp, UNCLASSIFIED, available from DTIC (AD-F860 122).

Helmbold, Robert L., Do Battles and Wars Have a Common Relationship Between Casualties and Victory?, US Army Concepts Analysis Agency, US Army Concepts Analysis Agency Technical Paper CAA-TP-87-16, November 1987, 52 pp, UNCLASSIFIED, available from DTIC (AD-A196 126).

Helmbold, Robert L., Personnel Attrition Rates in Historical Land Combat Operations: A Catalog of Attrition and Casualty Data Bases on Diskettes Usable With Personal Computers, CAA Research Paper CAA-RP-93-4, September 1993, AD-A279 069. "PAR Data Disks," (Diskettes accompanying the preceding research paper), AD-M000 344. Revised set of diskettes, AD-M000 368.

Helmbold, Robert L., "The Constant Fallacy: A Persistent Logical Flaw in Applications of Lanchester's Equations," European Journal of Operations Research, Vol 75 (1994), pp 647-658.

Kelley, C. T. Jr.; Smith, G. K.; and Matyskiela, W. W., "Quality-Quantity Tradeoffs: An Historical Analysis of Air Combat," The RAND Corporation, RAND Report R-1976-AF, April 1977, 45 pp, UNCLASSIFIED, available from DTIC (AD-C010 504).

Kirkpatrick, D. L. I., "Do Lanchester's Equations Adequately Model Real Battles?," RUSI (Journal of the Royal United Services Institute for Defence Studies), Vol 130, no 2, June 1985, pp 25-27, UNCLASSIFIED, available from libraries. See also the subsequent discussion of this paper in (i) Robertson, David, "Lanchester, the Civil War and Military Procurement," RUSI Vol 130, no 4, December 1985, pp 57-58, and (ii) Kirkpatrick, D. L. I., "Lanchester and Real Battles Revisited," RUSI Vol 130, no 4, December 1985, pp 59-60.

Laffin, John, Brassey's Battles: 3,500 Years of Conflict, Campaigns and Wars from A to Z, Brassey's Defence Publishers, London, ISBN: 0-08-031185-7, 1986, 484 pp, UNCLASSIFIED, available from publisher and libraries.

Lanchester, Frederick William, Aircraft in Warfare: The Dawn of the Fourth Arm, Constable & Co., London, 1916. UNCLASSIFIED, available from libraries. The portion dealing with Lanchester's square law was reprinted in The World of Mathematics, James R. Newman (ed.), Vol 4, pp 2138-2157, published by Simon & Schuster, New York, 1956.

Livermore, Thomas L., *Numbers and Losses in the Civil War in America, 1861-65*, Kraus Reprint Co., Milwood, NY, 1977 (a reprint of the Indiana University Press edition of 1957, originally published in Boston in 1900), 150 pp, UNCLASSIFIED, available from publisher and libraries.

Osipov, M., "The Influence of the Numerical Strength of Engaged Forces on Their Casualties," (in Russian), Military Collection, Numbers 6 through 10, 1915, (No 6, pp 59-74; No 7, pp 25-36; No 8, pp 31-40; No 9, pp 25-37; No 10, pp 93-96). UNCLASSIFIED, available from libraries. English translation published as US Army Concepts Analysis Research Paper CAA-RP-91-2, September 1991, and available from DTIC.

Peterson, Richard H., "On the 'Logarithmic Law' of Attrition and Its Application to Tank Combat," Operations Research, Vol 15, no 3, May-June 1967, pp 557-558, UNCLASSIFIED, available from libraries.

Richardson, Lewis Fry, *Statistics of Deadly Quarrels*, The Boxwood Press, Pacific Grove, California, ISBN 910286-10-8, 1960, 373 pp, UNCLASSIFIED, available from publisher and libraries. Edited by Quincy Wright and C. C. Lienau.

Rowland, David, "The Use of Historical Data in the Assessment of Combat Degradation," Journal of the Operational Research Society (UK), Vol 38, no 2, 1987, pp 149-162. UNCLASSIFIED, available from libraries. Also published as "Assessments of Combat Degradation" in the Journal of the Royal United Services Institute for Defence Studies (RUSI), June 1986, pp 33-43.

Smith, E. G.; and Donovan, G. N., "Trends in Warfare," British Army Operational Research Group, Memorandum No. F.6, April 1955, 38 pp, RESTRICTED, available from United Kingdom Defence Operational Analysis Establishment (DOAE), West Byfleet, UK.

Taylor, James G., *Lanchester Models of Warfare,* in two volumes, Military Applications Section, Operations Research Society of America, March 1983, UNCLASSIFIED, available from libraries.

Weiss, Herbert K., "Combat Models and Historical Data: The US Civil War," Journal of the Operations Research Society of America, Vol 14, Sep-Oct 1966, pp 759-790, UNCLASSIFIED, available from libraries.

Willard, Daniel, "Lanchester as Force in History," Research Analysis Corporation (RAC), Technical Paper RAC-TP-74, November 1962, 37 pp, UNCLASSIFIED, available from DTIC (AD-297 375L).

APPENDIX D

SOME OBSERVATIONS ON REGRESSION MODELS

D-1. ON THE COEFFICIENTS OF THE DIFFERENCE BETWEEN TWO REGRESSION MODELS

a. Consider the linear multivariate regression model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

where $\mathbf{y}$ is a known (N×1) matrix of observed quantities, $\mathbf{X}$ is a known (N×k) matrix of regressors or "independent" variables, β is a (k×1) matrix of unknown coefficients, and ε is an (N×1) matrix of residuals whose expected value is zero. It is well known that the ordinary least squares regression equations for the unknown parameters can be written in matrix form as

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y},$$

where $\mathbf{X}^T$ is the transpose of $\mathbf{X}$. The expected value of $\hat{\beta}$ is

$$E(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\, E(\mathbf{y}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta \equiv \beta.$$

b. Now suppose that we have two regressions to deal with. That is, suppose that

$$\mathbf{y}_1 = \mathbf{X}_1\beta_1 + \varepsilon_1$$
$$\mathbf{y}_2 = \mathbf{X}_2\beta_2 + \varepsilon_2$$

where the design matrices ($\mathbf{X}$), the residuals (ε), and the underlying regression coefficients (β) may differ. However, we do assume that the regression coefficient vectors β_1 and β_2 both have the same dimension, so that it makes sense to add and subtract them vectorially.

From the foregoing paragraphs, the estimated regression coefficients from these two regressions will be

$$\hat{\beta}_1 = (\mathbf{X}_1^T\mathbf{X}_1)^{-1}\mathbf{X}_1^T\mathbf{y}_1$$
$$\hat{\beta}_2 = (\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T\mathbf{y}_2$$

where $E(\hat{\beta}_1) = \beta_1$ and $E(\hat{\beta}_2) = \beta_2$. Now put

$$\mathbf{y}_0 = \mathbf{y}_2 - \mathbf{y}_1 = \mathbf{X}_0\beta_0 + \varepsilon_0$$

where

$$X_0 \beta_0 = X_2 \beta_2 - X_1 \beta_1$$
$$\varepsilon_0 = \varepsilon_2 - \varepsilon_1$$

and y_0 is defined for the set of values common to the y_2 and y_1 regressions. The question we raise is, "What can be said about how the regression coefficient β_0 is related to the regression coefficients β_2 and β_1"?

c. We break this question into two cases. In the first case, the design matrices are the same, that is, $X_2 = X_1 = X_0$. In this case,

$$\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y_0 = (X_0^T X_0)^{-1} X_0^T \{X_0 \beta_2 + \varepsilon_2 - (X_0 \beta_1 + \varepsilon_1)\} = \hat{\beta}_2 - \hat{\beta}_1$$

Hence, in this case, the relationship is simple and just what one might expect.

d. Now consider the case where the design matrices are not the same. Then we have

$$\hat{\beta}_0 = (X_0^T X_0)^{-1} X_0^T y_0 = (X_0^T X_0)^{-1} X_0^T \{X_2 \beta_2 - X_1 \beta_1 + \varepsilon_2 - \varepsilon_1\}$$

and, in this case, there is no simple general relationship between the βs. In fact, counterexamples can be created to demonstrate that, in this case, it is not necessarily true that $\beta_0 = \beta_2 - \beta_1$.

e. Accordingly, we can expect that $\beta_0 = \beta_2 - \beta_1$ only when the design matrices are identical, even in the OLS regression situation. The use of an iteratively reweighted regression robust method will introduce additional complications that may cause β_0 to fail to be equal to the difference between β_2 and β_1. Hence, in general, in this paper we do not expect to find that β_0 is given by the difference between β_2 and β_1.

D-2. SOLUTION OF A DIFFERENTIAL EQUATION MODEL

a. This paragraph is devoted to a solution of a differential equation model that explicitly takes into account the phenomenon of diminishing returns to scale. Our treatment here expands on the original treatment in Helmbold-1965. We begin with the usual Lanchester square law differential equations, which we write as

$$dx = -\beta y \, dt$$
$$dy = -\alpha x \, dt$$

Next, we assume that the attrition coefficients, α and β, depend on the instantaneous force ratio according to the following equations, which express one version of the phenomenon of diminishing returns to scale

$$\beta = B\left(\frac{y}{x}\right)^{w-1}$$

$$\alpha = A\left(\frac{x}{y}\right)^{w-1}$$

This leads to the following set of coupled nonlinear ordinary differential equations relating casualty rates to strengths, durations, and dates:

$$dx = -By(x/y)^h f(t)\,dt = -By^w x^{1-w} f(t)\,dt$$
$$dy = -Ax(y/x)^h f(t)\,dt = -Ax^w y^{1-w} f(t)\,dt$$

(D-2.1)

where $w = 1 - h$ is known as the Weiss parameter in honor of the famous military operations research analyst Herbert K. Weiss (see Taylor-1983).

b. We can separate variables via

$$\frac{dx}{dy} = \frac{B}{A}\frac{y^{2w-1}}{x^{2w-1}},$$

(D-2.2)

which (when $w \ne 0$) leads to the first integral

$$\frac{B}{A} = \frac{x_0^{2w} - x^{2w}}{y_0^{2w} - y^{2w}},$$

(D-2.3)

or (when $w = 0$) to the first integral

$$\frac{B}{A} = \frac{\ln(x/x_0)}{\ln(y/y_0)}.$$

(D-2.3a)

Accordingly, we name the following cases. If $w = 1$, we call (D-2.3) the Lanchester square law. If $w = 0$, we call (D-2.3a) the Peterson or logarithmic law. Note that (D-2.3a) is identical to the limiting value of (D-2.3) as w approaches 0. If $w = 1/2$, we call (D-2.3) the 1/2-linear law case. If $w = 1/4$, we call (D-2.3) the Osipov law case.

c. Also, note that the casualties to side X are approximately

$$C_x = x_0 - x \approx x_0 - \left(x_0 + \left.\frac{dx}{dt}\right|_0 T\right) = By_o^w x_0^{1-w} f(0)T,$$

where T is the time at the end of the battle, zero subscripts denote values at the start of the battle, and x or y subscripts denote the side. Similarly,

$$C_y \approx Ax_0^w y_0^{1-w} f(0)T.$$

Thus, the casualty exchange ratio favoring side Y is approximately

$$CERY \equiv \frac{C_x}{C_y} \approx \frac{B}{A}\left(\frac{y_0}{x_0}\right)^{2w-1}.$$

(D-2.4)

Define the (initial) force ratio favoring side Y to be $FRY = y_0 / x_0$, so that the fractional exchange ratio favoring side Y is

$$FERY \equiv \frac{C_x / x_0}{C_y / y_0} \equiv CERY\left(\frac{y_0}{x_0}\right) \equiv (CERY)(FRY) \approx \frac{B}{A}\left(\frac{y_0}{x_0}\right)^{2w}.$$

(D-2.5)

Later, we will define an advantage parameter favoring side Y, $ADVY$, which will turn out to be related to $FERY$ via

$$ADVY \approx \frac{1}{2}\ln(FERY) \approx \frac{1}{2}\ln\left\{\frac{B}{A}\left(\frac{y_0}{x_0}\right)^{2w}\right\} \approx \frac{1}{2}\ln\left(\frac{B}{A}\right) + w\ln\left(\frac{y_0}{x_0}\right).$$

(D-2.6)

d. Now put $\tau = g(t)$, where $g(\cdot)$ is at our disposal. Then

$$\frac{d\tau}{dt} = \frac{dg(t)}{dt}.$$

Changing variables from t to τ and denoting differentiation with respect to τ by primes,

$$x' = \frac{dx}{d\tau} = \frac{dx}{dt}\frac{dt}{d\tau} = -By^w x^{1-w} f(t) / \left(\frac{dg(t)}{dt}\right).$$

We now choose

$$\tau = g(t) = \int_{t_0}^{t} f(s)\,ds,$$

(D-2.7)

so that

$$x' = -By^w x^{1-w}$$
$$y' = -Ax^w y^{1-w}$$

(D-2.8)

in which x and y are to be construed as functions of τ, rather than of t, where τ is the integral of $f(t)$ given by (D-2.7). For example, if $f(t) = (\eta + \theta t)^c$ and $c \neq -1$, then

$$\tau = g(t) = \frac{1}{\theta b}\left\{(\eta + \theta t)^b + (\eta + \theta t_0)^b\right\},$$

where $b = c + 1$. If $c = -1$, then

$$\tau = \frac{1}{\theta} \ln\left(\frac{\eta + \theta t}{\eta + \theta t_0}\right).$$

e. Now put

$$a = \left(\frac{x}{x_0}\right)^k, \quad x = x_0 a^{1/k}$$

$$d = \left(\frac{y}{y_0}\right)^k, \quad y = y_0 d^{1/k}$$

(D-2.9)

where k is at our disposal. Then

$$x_0\left(\frac{1}{k}\right)a^{\frac{1-k}{k}} a' = x' = -By_0^w d^{w/k} x_0^{1-w} a^{\frac{1-w}{k}},$$

or

$$a' = -Bk\left(\frac{y_0}{x_0}\right)^w d^{w/k} a^{(k-w)/k}.$$

Now take $k = w$ so that

$$a' = -Bw\left(\frac{y_0}{x_0}\right)^w d = -\delta d.$$

(D-2.10a)

Similarly,

$$d' = -Aw\left(\frac{y_0}{x_0}\right)^{-w} a = -\alpha a,$$

(D-2.10b)

where we have put

$$\delta \equiv Bw\left(\frac{y_0}{x_0}\right)^w$$

$$\alpha \equiv Aw\left(\frac{y_0}{x_0}\right)^{-w}$$

(D-2.11)

We note that a first integral can be found as follows. Because

$$\frac{da}{dd} \equiv \frac{a'}{d'} = \frac{\delta}{\alpha}\frac{d}{a},$$

it follows that

$$\mu^2 \equiv \frac{\delta}{\alpha} = \frac{1-a^2}{1-d^2}. \qquad \text{(D-2.12)}$$

f. It is well known that the solution to (D-2.10) can be written as

$$a = \cosh(\lambda\tau) - \mu\sinh(\lambda\tau)$$
$$d = \cosh(\lambda\tau) - \frac{1}{\mu}\sinh(\lambda\tau) \qquad \text{(D-2.13)}$$

where we have put

$$\mu = \sqrt{\frac{\delta}{\alpha}} \equiv \left(\frac{y_0}{x_0}\right)^w \sqrt{\frac{B}{A}} \qquad \text{(D-2.14)}$$
$$\lambda = \sqrt{\alpha\delta} \equiv w\sqrt{AB}$$

Note that

$$\lambda\mu = \delta \equiv Bw\left(\frac{y_0}{x_0}\right)^w$$
$$\frac{\lambda}{\mu} = \alpha \equiv Aw\left(\frac{y_0}{x_0}\right)^{-w} \qquad \text{(D-2.15)}$$

g. Note that, with $C_z = z_0 - z$ being side Z's casualty number and $f_z = C_z / z_0$ being its casualty fraction, for small casualty fractions we have approximately

$$1 - a \approx 1 - \left(\frac{x}{x_0}\right)^w = 1 - \left(\frac{x_0 - C_x}{x_0}\right)^w = 1 - (1 - f_x)^w \approx wf_x.$$

Similarly, $1 - d \approx wf_y$. Note also that, for small values of $\varepsilon = \lambda\tau = \lambda g(t)$, expansion of the hyperbolic functions leads to the approximations

$$1 - a \approx 1 - (1 - \mu\lambda\tau) = \mu\lambda\tau = \delta\tau$$
$$1 - d \approx 1 - (1 - \frac{\lambda}{\mu}\tau) = \frac{\lambda}{\mu}\tau = \alpha\tau$$

Therefore, we have that, approximately

$$\delta\tau \approx wf_x$$
$$\alpha\tau \approx wf_y$$

Consequently,

$$\mu^2 \equiv \frac{\delta}{\alpha} \approx \frac{f_x}{f_y} \equiv FERY.$$

Also, $\lambda^2\tau^2 = \alpha\delta\tau^2 \approx w^2 f_x f_y$, so that

$$\varepsilon = \lambda\tau \approx w\sqrt{f_x f_y} \,.$$

h. More accurately, we have the usual inverse solution for the parameters:

$$\mu^2 = \frac{1-a^2}{1-d^2}$$
$$\varepsilon = \lambda\tau = \ln\left(\frac{1+\mu}{a+d\mu}\right)$$

Accordingly, the procedure for solving the inverse problem for μ and ε, given x_0, x, y_0, y, and w, is as follows. Compute

$$a = \left(\frac{x}{x_0}\right)^w$$
$$d = \left(\frac{y}{y_0}\right)^w$$
$$\mu^2 = \frac{1-a^2}{1-d^2}$$
$$\varepsilon = \ln\left(\frac{1+\mu}{a+d\mu}\right)$$

We may also find it handy to compute the advantage parameter favoring side Y as $ADVY \equiv \ln\mu = \frac{1}{2}\ln(\mu^2)$. If, in addition, we know $f(\cdot)$ or $g(\cdot)$ and T, then we can compute

$$\tau = g(T) = \int_0^T f(s)ds$$
$$\lambda = \varepsilon / \tau$$

i. Given x_0, y_0, A, B, $w \neq 0$, $g(\cdot)$, and t, the forward solution is as follows. (The modifications needed for the case $w = 0$ are obvious.) Put

$$\delta = Bw\left(\frac{y_0}{x_0}\right)^w$$

$$\alpha = Aw\left(\frac{y_0}{x_0}\right)^{-w}$$

$$\tau = g(t) = \int_{t_0}^{t} f(s)ds$$

$$\lambda = \sqrt{\alpha\delta}$$

$$\mu = \sqrt{\frac{\delta}{\alpha}}$$

$$\varepsilon = \lambda\tau$$

$$a(\tau) = \cosh\varepsilon - \mu\sinh\varepsilon$$

$$d(\tau) = \cosh\varepsilon - \frac{1}{\mu}\sinh\varepsilon$$

$$x(t) = x_0 a(\tau)^{1/w}$$

$$y(t) = y_0 d(\tau)^{1/w}$$

D-3. TABLES OF REGRESSION COEFFICIENT RESULTS

a. This paragraph presents regression coefficient results in tabular form. The results in graphical form were presented in Chapter 4. The graphical form is well-suited to visual comparisons and contrasts, while the tabular form is best for exact numerical comparisons.

b. All of the regression results presented in this paragraph were generated using the Number Cruncher Statistical System (NCSS) and Andrew's Sine robust regression with a tuning constant of 2.1, iterated until the change to the RMSE or to any of the regression coefficients was less than one percent.

Data Base	Side (A/D vs W/L)	Duration Type	Dependent Variable	Const Term	Const Std Err	Coeff lnX0	lnX0 Std Err	Coeff lnY0	lnY0 Std Err	Coeff T	T Std Err	Coeff Date Cent	Date Cent Std Err	DOF	RMSE	R-Sq
CDB91DAT	A/D	Active	lnCERY	-0.717	0.523	0.419	0.095	-0.362	0.084	0.059	0.318	-0.066	0.074	224	0.766	0.088
PARCOMBO	A/D	Active	lnCERY	-0.637	0.484	0.319	0.088	-0.251	0.072	-0.006	0.006	-0.012	0.080	130	0.477	0.100
CDB91DAT	A/D	Elapsed	lnCERY	-0.608	0.521	0.407	0.094	-0.365	0.085	0.131	0.160	-0.085	0.073	220	0.774	0.089
CDB91DAT	A/D	Overall	lnCERY	-0.514	0.427	0.381	0.085	-0.348	0.076	-0.032	0.041	-0.026	0.059	286	0.768	0.076
PARCOMBO	A/D	Overall	lnCERY	-1.131	0.295	0.482	0.054	-0.381	0.048	-0.005	0.002	-0.014	0.017	358	0.558	0.187
CDB91DAT	A/D	Active	lnCx	-0.464	0.368	0.689	0.069	0.144	0.065	-0.291	0.230	-0.596	0.053	224	0.553	0.804
PARCOMBO	A/D	Active	lnCx	-1.638	0.483	0.710	0.085	0.258	0.069	0.005	0.006	-0.154	0.076	130	0.501	0.794
CDB91DAT	A/D	Elapsed	lnCx	-0.430	0.386	0.644	0.074	0.172	0.069	0.212	0.137	-0.637	0.056	220	0.593	0.792
CDB91DAT	A/D	Overall	lnCx	-0.794	0.314	0.670	0.065	0.176	0.060	0.085	0.032	-0.557	0.044	287	0.582	0.792
PARCOMBO	A/D	Overall	lnCx	-2.327	0.268	0.763	0.048	0.251	0.042	0.009	0.002	-0.035	0.017	360	0.543	0.813
CDB91DAT	A/D	Active	lnCy	-0.954	0.413	0.408	0.075	0.495	0.068	-0.148	0.248	-0.519	0.060	225	0.614	0.796
PARCOMBO	A/D	Active	lnCy	-0.516	0.457	0.099	0.074	0.772	0.067	0.008	0.006	-0.430	0.087	132	0.530	0.799
CDB91DAT	A/D	Elapsed	lnCy	-0.629	0.411	0.393	0.075	0.464	0.068	0.246	0.133	-0.609	0.060	221	0.619	0.796
CDB91DAT	A/D	Overall	lnCy	-0.815	0.351	0.439	0.069	0.435	0.063	0.063	0.036	-0.572	0.048	288	0.642	0.778
PARCOMBO	A/D	Overall	lnCy	-0.956	0.277	0.278	0.049	0.616	0.043	0.017	0.003	-0.017	0.019	361	0.597	0.778
CDB91DAT	A/D	Active	lnFERY	-0.717	0.523	-0.581	0.095	0.638	0.084	0.059	0.318	-0.066	0.074	224	0.766	0.244
PARCOMBO	A/D	Active	lnFERY	-0.637	0.484	-0.681	0.088	0.749	0.072	-0.006	0.006	-0.012	0.080	130	0.477	0.495
CDB91DAT	A/D	Elapsed	lnFERY	-0.608	0.521	-0.593	0.094	0.635	0.085	0.131	0.160	-0.085	0.073	220	0.774	0.249
CDB91DAT	A/D	Overall	lnFERY	-0.514	0.427	-0.619	0.085	0.652	0.076	-0.032	0.041	-0.026	0.059	286	0.768	0.231
PARCOMBO	A/D	Overall	lnFERY	-1.131	0.295	-0.518	0.054	0.619	0.048	-0.005	0.002	-0.014	0.017	358	0.558	0.352

Table D-1. Table of Regression Coefficients for the Regressions Using Attacker and Defender Sides

Data Base	Side (A/D vs W/L)	Duration Type	Dependent Variable	Const Term	Const Std Err	Coeff lnX0	lnX0 Std Err	Coeff lnY0	lnY0 Std Err	Coeff T	T Std Err	Coeff Date Cent	Date Cent Std Err	DOF	RMSE	R-Sq
CDB91DAT	W/L	Active	lnCERW	1.610	0.362	-0.549	0.060	0.480	0.056	-0.684	0.226	0.077	0.050	222	0.527	0.320
PARCOMBO	W/L	Active	lnCERW	3.199	0.375	-0.496	0.053	0.237	0.057	-0.001	0.004	-0.052	0.056	130	0.360	0.488
CDB91DAT	W/L	Elapsed	lnCERW	1.821	0.354	-0.552	0.059	0.448	0.055	-0.192	0.110	0.024	0.049	218	0.516	0.310
BODASHIP	W/L	Overall	lnCERW	2.757	0.832	-0.327	0.155	0.232	0.147	-0.199	0.102	0.196	0.098	94	0.551	0.174
BWSHALL	W/L	Overall	lnCERW	4.473	0.245	-0.496	0.038	0.140	0.034	-0.001	0.004	-0.136	0.033	971	0.558	0.259
CDB91DAT	W/L	Overall	lnCERW	1.678	0.298	-0.552	0.053	0.482	0.049	-0.001	0.029	-0.045	0.039	284	0.524	0.300
PARCOMBO	W/L	Overall	lnCERW	2.205	0.214	-0.564	0.035	0.393	0.035	0.002	0.002	0.002	0.013	354	0.428	0.441
CDB91DAT	W/L	Active	lnCL	-0.027	0.306	0.412	0.052	0.425	0.048	-0.336	0.190	-0.525	0.044	225	0.468	0.854
PARCOMBO	W/L	Active	lnCL	0.469	0.405	0.325	0.060	0.485	0.064	0.005	0.006	-0.144	0.075	131	0.494	0.771
CDB91DAT	W/L	Elapsed	lnCL	0.155	0.299	0.409	0.051	0.390	0.048	0.341	0.102	-0.598	0.043	221	0.461	0.861
BODASHIP	W/L	Overall	lnCL	4.385	0.757	0.134	0.141	0.210	0.144	0.113	0.093	-0.244	0.094	97	0.508	0.341
BWSHALL	W/L	Overall	lnCL	1.418	0.182	0.175	0.026	0.502	0.024	0.011	0.003	-0.319	0.024	970	0.415	0.612
CDB91DAT	W/L	Overall	lnCL	-0.067	0.269	0.423	0.049	0.400	0.046	0.058	0.030	-0.585	0.036	289	0.503	0.836
PARCOMBO	W/L	Overall	lnCL	-0.170	0.215	0.346	0.035	0.490	0.035	0.018	0.002	-0.018	0.014	354	0.461	0.823
CDB91DAT	W/L	Active	lnCW	-1.578	0.424	0.804	0.073	0.077	0.070	0.693	0.284	-0.535	0.061	225	0.659	0.776
PARCOMBO	W/L	Active	lnCW	-2.149	0.483	0.926	0.070	0.066	0.070	0.007	0.006	-0.259	0.083	131	0.557	0.806
CDB91DAT	W/L	Elapsed	lnCW	-1.566	0.410	0.809	0.071	0.070	0.068	0.622	0.142	-0.572	0.059	221	0.638	0.791
BODASHIP	W/L	Overall	lnCW	0.359	0.597	0.402	0.111	0.190	0.108	0.320	0.069	-0.279	0.069	94	0.386	0.680
BWSHALL	W/L	Overall	lnCW	-3.138	0.239	0.674	0.034	0.387	0.033	0.011	0.004	-0.195	0.033	971	0.555	0.658
CDB91DAT	W/L	Overall	lnCW	-1.979	0.335	0.831	0.062	0.096	0.059	0.093	0.035	-0.399	0.046	287	0.626	0.783
PARCOMBO	W/L	Overall	lnCW	-2.731	0.253	0.851	0.042	0.183	0.043	0.010	0.002	-0.039	0.018	358	0.551	0.840
CDB91DAT	W/L	Active	lnFERW	1.610	0.362	0.451	0.060	-0.520	0.056	-0.684	0.226	0.077	0.050	222	0.527	0.358
PARCOMBO	W/L	Active	lnFERW	3.199	0.375	0.504	0.053	-0.783	0.057	-0.001	0.004	-0.052	0.056	130	0.360	0.591
CDB91DAT	W/L	Elapsed	lnFERW	1.821	0.354	0.448	0.059	-0.552	0.055	-0.192	0.110	0.024	0.049	218	0.516	0.364
BODASHIP	W/L	Overall	lnFERW	2.700	0.843	0.683	0.157	-0.773	0.150	-0.197	0.103	0.195	0.099	94	0.561	0.290
BWSHALL	W/L	Overall	lnFERW	4.472	0.245	0.503	0.038	-0.858	0.034	-0.001	0.004	-0.135	0.033	971	0.559	0.419
CDB91DAT	W/L	Overall	lnFERW	1.678	0.298	0.448	0.053	-0.538	0.049	-0.001	0.029	-0.045	0.039	284	0.524	0.309
PARCOMBO	W/L	Overall	lnFERW	2.205	0.214	0.436	0.035	-0.607	0.035	0.002	0.002	0.002	0.013	354	0.428	0.470

Table D-2. Table of Regression Coefficients for the Regressions Using Winner and Loser Sides

APPENDIX E

DISTRIBUTION

Addressee	No of copies
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZXA Washington, DC 20310	1
Deputy Chief of Staff for Logistics Headquarters, Department of the Army ATTN: DALO-ZXA-A Room 3D572, The Pentagon Washington, DC 20310-0580	1
Commander Combined Arms Support Command ATTN: ATCL-CFC Fort Lee, VA 23801-6000	1
Office of the Secretary of the Army Correspondence & Records Center Management Systems & Support ATTN: JDMSS-CRC ROOM 3D718, The Pentagon Washington, DC 20310-0105	1
Office of The Surgeon General ATTN: DASG-HCD 5109 Leesburg Pike Falls Church, VA 22041-3258	1
Director US Army TRADOC Analysis Command-WSMR ATTN: ATRC-WSL White Sands Missile Range, NM 88002-5502	1
Director, TRAC ATTN: ATRC-TD Fort Leavenworth, KS 66027-5200	1

Addressee	No of copies
HQ TRADOC ATTN: ATAN-S Fort Monroe, VA 23651-5143	1
Director US Army Materiel Systems Analysis Activity ATTN: AMXSY-LM Aberdeen Proving Ground, MD 21005-5071	1
Director US Army Research Laboratory ATTN: SARL-DD-T Building 305 Aberdeen Proving Ground, MD 21005-5066	1
Commander USACAC, CD ATTN: ATZL-CDE Fort Leavenworth, KS 66027-5300	1
Commander US Army Test and Evaluation Command ATTN: AMSTE-SI-S Aberdeen Proving Ground, MD 21005-5000	1
Commander Foreign Science and Technology Center 220 7th Street NE Charlottesville, VA 22901-5396	1
Director Defense Nuclear Agency ATTN: LASS 6801 Telegraph Road Alexandria, VA 22310-3398	1
Commander US Total Army Personnel Agency ATTN: DAPC-ZA 200 Stovall Street Alexandria, VA 22332	1

Addressee	No of copies
Commander US Army Quartermaster Center and School ATTN: ATSM-CES-0 Fort Lee, VA 23801	1
Defense Technical Information Center ATTN: DTIC-FDAC Cameron Station Alexandria, VA 22314-6145	2
US Army Nuclear and Chemical Agency 7500 Backlick Road, Bldg #2073 ATTN: MONA-AD/Library Springfield, VA 22150-3198	1
USASCAF The Pentagon Library ATTN: JDHQ-LR (Army Studies) Room 1A518, The Pentagon Washington, DC 20310-6000	1
CINC Forces Command Force Capabilities Div ATTN: FCJ5-FC Fort McPherson, GA 30330-6000	1
OSD (PA&E) (DC&L) Room 2E313, The Pentagon ATTN: Mr. J. Johnson Washington, DC 20310-1800	1
Integration and Assessment Division Joint Staff/J8 (LTC Ford) Room 1D964, The Pentagon Washington, DC 20318-8000	1
Joint Chiefs of Staff SJCS, Documents Division ATTN: RAIR Branch Room 2B939, The Pentagon Washington, DC 20310-5000	1

Addressee	No of copies
Commandant US Army War College Operations Group ATTN: AWCM-A Carlisle Barracks, PA 17013-5050	1
Air War College ATTN: AU/CADRE/WGOI Maxwell Air Force Base, AL 36112-5532	1
President US Navy War College ATTN: Code E-111/Library Newport, RI 02841-5010	1
President National Defense University ATTN: NDU-LD-CDC Washington, DC 20319-6000	1
Commandant Armed Forces Staff College ATTN: Library Room B-201 7800 Hampton Blvd Norfolk, VA 23511-6097	1
Commandant US Army Command and General Staff College ATTN: ATZL-SWS-L (Mail) Fort Leavenworth, KS 66027-6900	1
United States Military Academy ATTN: MAIM-SC-A West Point, NY 10996-5000	1
Superintendent Naval Postgraduate School ATTN: Security Manager Monterey, CA 93940	1

Addressee	No of copies
Commander US Army Ordnance Center and School ATTN: ATSL-CC Aberdeen Proving Ground, MD 21005-6500	1
Commander US Army, Pacific ATTN: APOP-SPM/Security Officer Fort Shafter, HI 96858-5100	1
Commander US Army Health Services Command ATTN: HSOP-FSI Fort Sam Houston, TX 78234-6000	1
Commander, USAITAC AIAIT-HI, Tech Info Bldg 203, STOP 314 Washington Navy Yard Washington, DC 20374	1
Commander/Director US Army Engineer Studies Center Casey Building, No. 2594 ATTN: ESC-AO (Security Officer) Fort Belvoir, VA 22060-5583	1
Commander US Army Missile Command ATTN: AMSMI-OR-SA Redstone Arsenal, AL 35898-5060	1
Commander US Army Training and Doctrine Command ATTN: ATIM-ATCD Fort Monroe, VA 23651-5000	1
Headquarters US Army Materiel Command ATTN: AMCPE-AR 5001 Eisenhower Avenue Alexandria, VA 22333-0001	1

Addressee	No of copies

Commandant
Air Force Institute of Technology
ATTN: AFIT-EN
Wright-Patterson AFB, OH 45433

1

Commandant
US Marine Corps
ATTN: HQSR-3
Washington, DC 20380

1

Internal Distribution:
Reference copy:
Unclassified Library

2

Record copy:
Originating office (CSCA-TCT)

1

Other copies:
Dr. Robert L. Helmbold

10

GLOSSARY

1. INTRODUCTION. Some of the abbreviations and special terms used in this document are listed below. If the definition given is an official one, the organizations that have adopted it are given in parentheses; otherwise, no indication of its adoption are given. Note that the definitions used by other countries or by the US in earlier times may differ more or less from those given below, and may be interpreted in various ways even within the US Department of Defense.

2. DEFINITION OF TERMS

Battle casualty - (DOD) Any casualty incurred in action. "In action" characterizes the casualty status as having been the direct result of hostile action, sustained in combat or relating thereto, or sustained going to or returning from a combat mission provided that the occurrence was directly related to hostile action. Included are persons killed or wounded mistakenly or accidentally by friendly fire directed at a hostile force or what is thought to be a hostile force. However, not to be considered as sustained in action and thereby not to be interpreted as battle casualties are injuries due to the elements, self-inflicted wounds, and, except in unusual cases, wounds or death inflicted by friendly forces while the individual is in absent without leave or dropped from rolls status or is voluntarily absent from a place of duty. See also died of wounds received in action; nonbattle casualty; wounded.

Bloody losses - The sum of the KIA and WIA.

Casualty - (DOD, IADB) Any person who is lost to the organization by reason of having been declared dead, wounded, injured, diseased, interned, captured, retained, missing, missing in action, beleaguered, besieged or detained; see also battle casualty; nonbattle casualty; wounded.

CMIA - Captured or missing in action. See POW and MIA.

CRO - Carded for record only. (Adapted from Beebe, Gilbert W.; and De Bakey, Michael E., *Battle Casualties: Incidence, Mortality, and Logistic Considerations,* Charles C. Thomas (publisher), 1952.) Basically, admissions to a medical treatment facility include all cases admitted for medical care and not returned to duty on the same calendar day as that on which first seen. Cases which are treated on an outpatient (duty) status, are designated as carded for record only (CRO).

DNBI - Disease and nonbattle injury. Personnel treated for diseases and for injuries not received in action. See Nonbattle casualty.

DOW - Died of wounds received in action (DOD, NATO). A battle casualty who dies of wounds or other injuries received in action, after having reached a medical treatment facility. See also killed in action.

DTIC - Defense Technical Information Center.

KIA - Killed in action (DOD, NATO, IADB). A battle casualty who is killed outright or who dies as a result of wounds or other injuries before reaching a medical treatment facility. See also died of wounds received in action.

Losses - (Adapted from FM 101-10-1/2, Staff Officers' Field Manual Organizational, Technical, and Logistical Data Planning Factors, October 1987). A personnel loss is any reduction in the assigned strength of a unit. Personnel losses are recorded in three general categories: battle, nonbattle, and administrative.

- Battle losses are those incurred in action. They include wounded or injured in action (including those who died of wounds and died of injuries received in action), killed in action, and missing in action or captured by the enemy.

- Nonbattle losses are those not directly attributable to action regardless of when sustained. They include nonbattle dead, nonbattle accident/injury, nonbattle missing, and illness/disease.

- Administrative losses are those resulting from transfer from the unit, absence without leave, desertion, personnel rotation, and discharges.

LWIA - Lightly wounded in action (see Slightly wounded).

MIA - (adapted from FM 101-10-1/2, Staff Officers' Field Manual Organizational, Technical, and Logistical Data Planning Factors, October 1987). Missing in action describes battle casualties whose whereabouts or fate cannot be determined and who are not known to be in an unauthorized absence status (desertion or absence without leave). Missing in action (MIA) casualties are not usually included in medical statistical records or reports received by The Surgeon General, but are reportable to The Adjutant General.

NFW - Nonfatal wound. A person who is wounded in action (WIA), but who does not die of wounds (DOW).

Nonbattle casualty - (DOD, NATO, IADB) A person who is not a battle casualty but who is lost to his organization by reason of disease or injury, including persons dying from disease or injury, or by reason of being missing where the absence does not appear to be voluntary or due to enemy action. See also battle casualty; wounded.

Nonbloody loss - Battle casualties other than KIA and WIA; include (for example) MIA, POW, absent without leave, stragglers, and deserters.

NP - Neuropsychiatric.

POW - Prisoner of war. Detainee (DOD). A term used to refer to any person captured or otherwise detained by an armed force. (According to FM 101-10-1/2, Staff Officers' Field Manual Organizational, Technical, and Logistical Data Planning Factors, October 1987, captured describes all battle casualties known to have been taken into custody by a hostile force as a result of and for reasons arising out of any armed conflict in which US armed forces are engaged.

Captured casualties are not usually included in medical statistical records or reports received by The Surgeon General but are reported to The Adjutant General.)

Seriously wounded - (DOD, IADB) A stretcher case. See also WIA.

Slightly wounded - (DOD, IADB) A casualty that is a sitting or walking case. See also WIA.

SWIA - Seriously wounded in action (see Seriously wounded).

TBC - Total battle casualty. The sum of the KIA, WIA, and CMIA casualties.

WIA - Wounded in action (DOD, NATO, IADB). A battle casualty other than "killed in action" who has incurred an injury due to an external agent or cause. The term encompasses all kinds of wounds and other injuries incurred in action, whether there is a piercing of the body, as in a penetrating or perforated wound, or none, as in the contused wound; all fractures, burns, blast concussions, all effects of biological and chemical warfare agents, the effects of exposure to ionizing radiation, or any other destructive weapon or agent.

/kpd - Used as an abbreviation for the phrase "per thousand per day." Thus, the statement that "the attrition rate amounted to 10 per thousand per day" is abbreviated to "the attrition rate amounted to 10/kpd."

3. TERMS AND MATHEMATICAL SYMBOLS UNIQUE TO THIS STUDY. For more detail on notation, see paragraph 4-2 of the main body.

Force identification, Z: the symbol Z will identify the force under consideration. For example, Z may stand for X, Y, W, or L, which identify the following (respectively): side X–the attacker, side Y–the defender, side W–the winner, or side L–the loser. Side Z's opponent will be identified by the symbol $\overline{Z}$. Thus,

$$\overline{Z} = \begin{cases} Y \text{ if } Z = X \\ X \text{ if } Z = Y \\ L \text{ if } Z = W \\ W \text{ if } Z = L \end{cases}$$

Variable time, t : the symbol t will stand for a variable time, which increases from 0 at the start of the battle to the total battle time T at its end. In all cases, if the time t is not specified, then it is understood to be equal to the total engagement duration, T.

(Surviving personnel) strength, $Z(t)$ at time t : the personnel strength of side Z as a function of time t into the action.

(Initial personnel) strength, $Z0 = Z(0)$: the initial personnel strength of side Z. The symbols X and Y will always be used to refer to the attacking force and the defending force,

respectively. Hence, $X0$ is the attacker's initial personnel strength and $Y0$ is the defender's initial personnel strength.

(Personnel battle) casualties, $CZ(t)$ at time t : the number of personnel casualties inflicted on side Z from the beginning of an action up to time t into the action. In the absence of exogenous changes in strength (such as reinforcements and detachments), we have the "mass balance" identity: $CZ(t) = Z0 - Z(t)$. Here, by "exogenous changes" we mean any changes that affect the (personnel) strength of a force, other than those resulting from battle casualties. In this and all similar cases, if the time t is not specified, then it is understood to be equal to the total engagement duration, T. For example, CZ is the number of personnel casualties inflicted on side Z during the entire course of the battle.

Casualty fraction, $FZ(t) = CZ(t) / Z0$ at time t : the ratio of the number of personnel casualties inflicted on side Z during time t to the unit's initial personnel strength.

Surviving fraction, $SZ(t) = Z(t) / Z0$ at time t : the surviving fraction of force Z, as of time t into the action. In the absence of exogenous changes in strength, we have the obvious identity, $FX(t) = 1 - Z(t) / Z0 \equiv 1 - SZ(t)$.

(Average) Casualty fraction rate at time t , $RZ(t) = FZ(t) / t$: the average rate of increase in the casualty fraction from time 0 to time t .

Differential casualty fraction rate at time t , $d\,FZ(t) / dt$: the derivative of the casualty fraction (or of an appropriately smoothed version of it), evaluated at time t .

Force ratio favoring side Z at time t , $Z(t) / \overline{Z}(t)$: the ratio of side Z's strength to that of its opponent, $\overline{Z}$, evaluated at time t .

(Average) Casualty exchange ratio favoring side Z at time t , $CERZ(t) = C\overline{Z}(t) / CZ(t)$: the ratio of the casualty number on side $\overline{Z}$ to that on side Z.

Differential casualty exchange ratio favoring side Z at time t , $DFERZ(t) = d\,C\overline{Z}(t) / d\,CZ(t)$: the ratio of side $\overline{Z}$'s to side Z's instantaneous or differential casualty number, evaluated at time t . (In some cases, these differentials may be estimated using some appropriately smoothed representation of the casualty number.)

(Average) Fractional exchange ratio favoring side Z at time t , $FERZ(t) = F\overline{Z}(t) / FZ(t)$: the ratio of side $\overline{Z}$'s to side Z's casualty fraction, evaluated at time t .

Differential fractional exchange ratio favoring side Z at time t , $DFERZ(t) = d\,F\overline{Z}(t) / d\,FZ(t)$: the ratio of side $\overline{Z}$'s to side Z's instantaneous casualty fraction, evaluated at time t .